

**Aplicación de Redes Neuronales Artificiales en la Modelación de la Deforestación Asociada a Nuevos Proyectos de Infraestructura Vial en las Regiones del Nordeste y Bajo Cauca del Departamento de Antioquia**

**Trabajo de Grado Para Optar al Título de Magister en Medio Ambiente y Desarrollo**

**Luisa Fernanda Gómez Ossa**

**Directora**

**Verónica Botero Fernández**



**Universidad Nacional de Colombia, Sede Medellín  
Facultad de Minas**

**Departamento de Geociencias y Medio Ambiente  
Medellín 14 de octubre de 2014**

## AGRADECIMIENTOS

Agradezco con toda el alma a mi madre y hermanos Julián y Carolina por el apoyo que me han brindado. Su amor, esfuerzo, constancia y disciplina han sido un ejemplo y han contribuido a mi formación personal y profesional.

Agradezco a mi profesora Verónica Botero por su confianza y disposición, sus conocimientos han sido fundamentales para mi desarrollo profesional y personal.

Agradezco al **Convenio ISA-Universidad Nacional** por otorgarme la beca para realizar mis estudios de posgrado, fue una gran felicidad y alivio para mí.

Por último quiero dar las gracias a todos aquellos que considero mis amigos, por compartir conmigo momentos buenos y malos y porque también hicieron parte de todo mi proceso de formación.

## RESUMEN

La deforestación tropical es un proceso continuo causado principalmente por la construcción de nuevas vías, las cuales sin una planificación ambiental adecuada contribuyen a la pérdida de biodiversidad. En la presente investigación se estimó un modelo de regresión logística y un modelo de redes neuronales artificiales (RNAs), para predecir la deforestación asociada a nuevas vías en las regiones del Bajo Cauca y Nordeste del departamento de Antioquia para el periodo 1980-2000. Se consideraron variables biofísicas como la pendiente, la aptitud agrícola y variables que representan accesibilidad a mercados como la distancia a ríos, distancia a vías, distancia de menor costo a las cabeceras municipales y distancia al borde del bosque. La regresión logística se realizó para identificar los principales determinantes de la deforestación y dado que las RNAs tienen la capacidad de capturar relaciones no lineales, el conjunto de variables que generó mejores predicciones a partir del modelo de regresión logística se utilizó para la construcción y entrenamiento de RNAs. El entrenamiento se realizó en línea (modo on line) con el algoritmo de retropropagación, en el software R. Para probar la capacidad de predicción de los modelos se evaluó el área bajo la curva ROC (AUC). El modelo de regresión logística presentó un AUC de 0.77 y las RNAs un AUC de 0.79 a 0.82. Con las redes que presentaron altos valores de AUC se realizó un ensamble, a partir del cual se estimó la superficie de deforestación para un escenario base y un escenario simulado incorporando nuevas vías como la variante Porce y la vía El Bagre-San Jacinto del Cauca. La regresión logística indica que los principales factores de la deforestación para el periodo 1980-2000 fueron la distancia a las vías y la distancia al borde del bosque. La superficie estimada por el ensamble de RNAs muestra que los bosques más susceptibles a la deforestación se encuentran cerca de los centros poblados y siguen la localización de las principales vías. Finalmente la comparación de escenarios indica que la construcción de las nuevas vías conduciría a una deforestación de aproximadamente 10,782 ha.

**Palabras claves:** Redes neuronales artificiales, regresión logística, predicción, deforestación.

## **ABSTRACT**

Tropical deforestation is a continuous process caused mainly by the construction of new roads, which without proper environmental planning contribute to biodiversity loss. In this research, logistic regression models and artificial neural networks (ANNs) were estimated to predict the deforestation associated with new roads in the regions of Bajo Cauca and Northeast of the department of Antioquia for the period 1980-2000. Biophysical variables such as slope, agricultural suitability and accessibility variables representing distance to markets, such as distance to rivers, distance to roads, lowest cost distance to the municipalities and distance to forest edge were considered. Logistic regression was performed to identify the main determinants of deforestation and since the ANNs have the ability to capture nonlinear relationships, the set of variables that generated better predictions from the logistic regression model was used for the construction and training ANNs. The training was conducted online (online mode) with the backpropagation algorithm in the software R. To test the predictive power of the models the area under the ROC curve (AUC) was evaluated. The logistic regression model showed an AUC of 0.77 and ANNs an AUC of 0.79 to 0.82. With the networks that showed high AUC an assembly was held, from which the deforestation surface for a baseline scenario and a simulated scenario that incorporates new roads as Porce variant and the Bagre-San Jacinto del Cauca route was estimated. Logistic regression indicated that the main drivers of deforestation for the period 1980-2000 were the distance to roads and distance to forest edge. The estimated assembly ANNs surface shows that the most susceptible forests to deforestation are located near population centers and follow the location of the main roads. Finally the comparison of both scenarios indicates that the construction of new roads could lead to deforestation of approximately 10,782 ha.

**Keywords:** Artificial neural networks, logistic regression, prediction, deforestation.

## Tabla de Contenido

	Pág.
1. INTRODUCCIÓN .....	8
2. OBJETIVOS .....	10
3. PREGUNTAS DE INVESTIGACIÓN .....	11
4. HIPÓTESIS .....	11
5. REVISIÓN DE LITERATURA .....	12
6. METODOLOGÍA.....	20
6.1 Área de Estudio .....	20
6.2 Datos y Variables.....	21
6.3 Creación de variables explicadoras de la deforestación .....	22
6.4 Creación de la variable dependiente .....	29
6.5 Tasa de deforestación .....	30
6.6 Modelación de la deforestación .....	31
6.6.1 Modelo de Regresión Logística .....	31
6.6.2 Modelo de Red Neuronal .....	32
6.7 Validación.....	41
6.8 Ensamble de Redes Neuronales .....	42
6.9 Estimación del Coeficiente de Kappa.....	43
6.10 Simulación del Efecto de la Construcción de Nuevas Vías. ....	44
7. RESULTADOS .....	47
8. DISCUSIÓN.....	64
9. CONCLUSIONES .....	71
10. FUTURAS INVESTIGACIONES .....	74
BIBLIOGRAFÍA .....	76

## Lista de Figuras

	Pág.
<b>Figura 1.</b> Localización del área de estudio correspondiente a las subregiones del Nordeste y Bajo Cauca del departamento de Antioquia.....	20
<b>Figura 2.</b> Distancia euclidiana a ríos. Tamaño de pixel 30m .....	23
<b>Figura 3.</b> Distancia euclidiana a vías. Tamaño de pixel 30m .....	23
<b>Figura 4.</b> Distancia al borde del bosque/No bosque. Tamaño de pixel 30m .....	24
<b>Figura 5.</b> Distancia de menor costo a las cabeceras municipales. Tamaño de pixel 30m ...	27
<b>Figura 6.</b> Pendiente (grados). Tamaño de pixel 30m .....	28
<b>Figura 7.</b> Aptitud Agrícola. Tamaño de pixel 30m .....	29
<b>Figura 8.</b> Variable dependiente. Tamaño de pixel 30m .....	30
<b>Figura 9.</b> Esquema de un Perceptrón Multicapa .....	34
<b>Figura 10.</b> Vías proyectadas para la región de estudio. ....	45
<b>Figura 11.</b> Red vial empleada para la construcción de las variables explicadoras.....	46
<b>Figura 12.</b> Coberturas Terrestres de las Regiones del Nordeste y Bajo Cauca, Antioquia ..	49
<b>Figura 13.</b> Deforestación observada en las regiones del Nordeste y Bajo Cauca, Antioquia, 1980-2000.....	50
<b>Figura 14.</b> Reducción del error para un ciclo de entrenamiento de 5,000 iteraciones .....	54
<b>Figura 15.</b> Comportamiento del error con respecto a las tasas de entrenamiento y al número de neuronas ocultas para un ciclo de 5000 iteraciones. ....	55
<b>Figura 16.</b> Área bajo la curva ROC (AUC), con respecto a las tasas de entrenamiento y al número de neuronas ocultas para un ciclo de 5000 iteraciones. ....	55
<b>Figura 17.</b> Evaluación de la capacidad predictiva entre el ensamble de RNAs y la Regresión Logística (ROC/AUC). ....	57
<b>Figura 18.</b> Superficie de probabilidad de deforestación para las regiones del Bajo Cauca y Nordeste, Antioquia. ....	58
<b>Figura 19.</b> Coeficiente de Kappa para diferentes umbrales.....	59
<b>Figura 20.</b> Deforestación estimada para las regiones del Bajo Cauca y Nordeste, Antioquia	60
<b>Figura 21.</b> Localización de las zonas que presentarían un aumento en la probabilidad de deforestación por la Construcción de la Variante Porce y la vía Bagre- San Jacinto del Cauca .....	62

<b>Figura 22.</b> Localización de la deforestación neta adicional por la construcción de la Variante Porce y la Vía Bagre- San Jacinto del Cauca.....	63
---	----

### Lista de Tablas

	<b>Pág.</b>
<b>Tabla 1.</b> Valoración del coeficiente kappa .....	43
<b>Tabla 2.</b> Matriz de cambios en coberturas terrestres, 1980-2000.....	48
<b>Tabla 3.</b> Coeficientes de regresión estimados para una muestra de 700.000 puntos. ....	51
<b>Tabla 4.</b> Coeficientes de regresión estimados para una muestra balanceada de 250.000. ....	52
<b>Tabla 6.</b> Parámetros de las redes neuronales artificiales que presentaron mejor capacidad predictiva. ....	56

### Lista de Anexos

	<b>Pág.</b>
<b>ANEXO A.</b> Correlación entre las variables explicadoras de la deforestación para diferentes tamaños de muestra.....	84
<b>ANEXO B.</b> Redes entrenadas con diferente conjunto de Parámetros. ....	86
<b>ANEXO C.</b> Código implementado en RStudio.....	87

## 1. INTRODUCCIÓN

Por sus condiciones biogeográficas Colombia posee una gran biodiversidad, pero desafortunadamente se encuentra entre los 10 primeros países que ha presentado pérdidas significativas de hábitat boscoso, con una tasa de destrucción del 0.5% anual (Savage et al., 2010). Se proyecta que para el año 2050 casi el 80% de extinción de especies sea resultado de la deforestación (Rodríguez et al., 2012), un proceso continuo en América latina impulsado por el desarrollo de proyectos de infraestructura vial (Geist & Lambin 2001), los cuales sin una planificación ambiental adecuada contribuyen a la fragmentación de los bosques y a la pérdida de biodiversidad.

Con respecto a los impactos generados por las vías, desde una perspectiva biológica, disminuyen la diversidad de especies que en función de sus características únicas, son vulnerables a los cambios ambientales asociados a las vías y claros lineales, y desde una perspectiva socioeconómica las vías facilitan la colonización del bosque y la expansión de caminos ilegales (Laurance et al., 2009), un problema constante por la débil aplicación de las leyes ambientales en muchos países tropicales.

Además del impacto inmediato de la construcción de una vía dentro o cerca de áreas boscosas, las vías son elementos permanentes del paisaje que facilitan mayor accesibilidad a los bosques. Incluso cuando las vías no influyen directamente en la deforestación, representan un determinante a largo plazo de la conversión de bosques a usos alternativos, especialmente en paisajes con una larga historia de ocupación humana (Freitas et al., 2010) y aunque en algunas regiones es difícil asignar una causalidad directa, en áreas remotas se podrían evaluar cambios en el paisaje causados por la construcción de una vía.

En los proyectos viales se realizan evaluaciones ambientales para mitigar el impacto sobre el medio ambiente, pero en general estas evaluaciones se realizan a nivel local sin tener en cuenta los efectos a largo plazo en el entorno natural. Además, las principales fallas



como el carácter descriptivo de las evaluaciones, la falta de indicadores mensurables y la ausencia de predicciones cuantitativas (Mallard & Francois, 2012) no permiten establecer con rigor los impactos potenciales de estos proyectos, lo cual podría comprometer la conservación de recursos valiosos en el largo plazo.

En los estudios de impacto ambiental, los temas más recurrentes suelen ser inventarios florísticos y faunísticos. Estos estudios se deberían complementar con análisis de modelos espaciales, los cuales además de proporcionar información valiosa sobre determinantes de conversión de bosques, también son útiles en la elaboración de mapas que permitan identificar la localización de aquellos bosques más susceptibles a la deforestación, lo cual contribuye a priorizar actividades de gestión y conservación.

En Colombia se han realizado trabajos sobre determinantes de la deforestación, principalmente en la Amazonia (Armenteras et al., 2006; Etter et al., 2006a; Etter et al., 2006c) con técnicas propias de la ecología del paisaje. En el departamento de Antioquia se realizó un estudio de deforestación (Orrego 2009) en el cual se usó teoría de la renta y un modelo logit para identificar las principales causas de la conversión de bosques. Aunque estas investigaciones han permitido identificar los lugares más susceptibles a la deforestación, no se ha profundizado en analizar o estimar la condición de los bosques luego de la construcción de nuevos proyectos de infraestructura vial, lo cual ayudaría a mejorar la gestión y conservación de los bosques a nivel local.

Modelos estadísticos han sido ampliamente utilizados para identificar los determinantes de la deforestación tropical. Además de estos métodos, existen modelos de simulación dinámica como las redes neuronales artificiales que tienen la capacidad de realizar predicciones ante la presencia de relaciones no lineales (Mas et al., 2004). En la presente investigación se estimó un modelo de regresión logística y un modelo con redes neuronales artificiales, para las regiones del Nordeste y Bajo Cauca del departamento de Antioquia. El modelo de regresión logística se desarrolló para identificar la importancia

relativa de las variables explicadoras y las RNAs para predecir la deforestación asociada a nuevos proyectos de infraestructura vial.

## **2. OBJETIVOS**

### **2.1 Objetivo General**

- Modelar a partir de Redes Neuronales Artificiales la deforestación asociada a nuevos proyectos de infraestructura vial en las regiones del Nordeste y Bajo Cauca del departamento de Antioquia.

### **2.2 Objetivos Específicos**

- Identificar los principales determinantes biofísicos de la deforestación en las regiones del Nordeste y Bajo Cauca del departamento de Antioquia para el período 1980-2000 a partir de un modelo de regresión logística.
- Comparar la capacidad predictiva de las Redes neuronales artificiales en la modelación de la deforestación con respecto al método de regresión logística.
- Simular con redes neuronales artificiales el efecto de la construcción de la Variante Porce y la vía El Bagre- San Jacinto del Cauca sobre la deforestación en las regiones del Nordeste y Bajo Cauca del departamento de Antioquia.

### **3. PREGUNTAS DE INVESTIGACIÓN**

- 1.** Cuáles son los principales determinantes biofísicos de la deforestación presente en el período 1980-2000 para las regiones del Nordeste y bajo Cauca del departamento de Antioquia?
- 2.** Las Redes Neuronales Artificiales tienen la capacidad de predecir la deforestación en las regiones del Nordeste y Bajo Cauca del departamento de Antioquia con una precisión superior al modelo de regresión logística?
- 3.**Cuál sería la deforestación adicional generada por la construcción de la variante Porce y la vía El Bagre- San Jacinto del cauca, en las regiones del Nordeste y Bajo Cauca del departamento de Antioquia?

### **4. HIPÓTESIS**

Las Redes Neuronales Artificiales tienen una capacidad predictiva superior a la regresión logística y podrían ser útiles para simular los efectos de la construcción de nuevas vías sobre la deforestación.

## 5. REVISIÓN DE LITERATURA

En las últimas cinco décadas los determinantes de naturaleza antrópica de la deforestación han cambiado sustancialmente. Esto tiene implicaciones importantes en los esfuerzos de conservación. En una investigación reciente, un meta-análisis de 268 casos de estudio, permitió identificar los determinantes sociales asociados con la deforestación tropical antes y después de 1990 (Rudel et al., 2009). Los principales resultados del estudio indican que pequeños agricultores contribuyeron a la deforestación de grandes extensiones de bosques tropicales antes de 1990, en el sureste de Asia y en América Latina. Después de 1990, la ganadería, la agricultura y las plantaciones se identificaron como los principales factores que influyeron en la deforestación en el sudeste asiático y en la cuenca del Amazonas.

Además de agentes antrópicos, variables ambientales y geográficas como la proximidad a vías y centros urbanos, son de vital importancia para evaluar el riesgo de conversión de los bosques a usos alternativos (Gaston et al., 2002). Por ejemplo, en la Amazonia brasileña, la distancia a vías principales fue un predictor significativo de la deforestación (Laurance et al., 2002). El bosque de Basho al norte de Pakistán se ha reducido en al menos un 50% debido a la extracción ilegal de madera, que se llevó a cabo luego de la apertura del valle a través de la construcción de una carretera de enlace en 1968 (Ali et al., 2005) y Cropper et al. (2001) en su estudio realizado en Tailandia predijeron la deforestación en las áreas protegidas debido a la construcción de nuevas vías.

La deforestación es un tema que ha recibido una importante atención durante los últimos años por las consecuencias negativas en la conservación de la biodiversidad (Etter et al., 2006a). Una característica notable de los casos reportados de deforestación es que no existe causalidad por variables individuales, sino por la combinación de factores sinérgicos que resultan de procesos socio-económicos complejos y en muchos casos es imposible aislar una única causa. Por ejemplo Geist & Lambin (2001) encontraron que la extensión de la infraestructura vial, en combinación con otras causas próximas, como la expansión

de la agricultura explicó 110 de 152 casos de deforestación reportados en la región tropical.

La mayoría de bosques han sido y continúan siendo fragmentados por la construcción, uso y mantenimiento de las vías, las cuales juegan un papel importante en determinar los patrones de deforestación. Su ubicación determina las condiciones, las facilidades y las distancias de extracción y transporte de productos agrícolas, y con ello, se establece una clara diferencia en los costos de producción (Vélez et al., 2010). Esto explica porqué las carreteras y las vías de penetración inducen la colonización a medida que se construyen o aún mucho antes de iniciar trabajos de explanación. Por ejemplo en la selva amazónica se identificó un patrón denominado espina de pescado causada por el desmonte del bosque en hileras, desde el borde de las vías hacia el bosque remanente para dar paso a pequeñas fincas agrícolas (Freitas et al., 2010). Igualmente en la Amazonia Brasileña el 95% de toda la deforestación se presentó a menos de 50 km de las autopistas (Laurance et al., 2009) y en Nueva Zelanda y Estados Unidos la alta densidad de vías indicó un uso intensivo del paisaje, por tanto las vías fueron un importante determinante de la deforestación (Bresee et al., 2004; Ewers et al., 2006).

Como se mencionó anteriormente son varias las investigaciones que han demostrado como la construcción de nuevas vías incrementa las tasas de deforestación por facilitar el acceso a las zonas boscosas. Para apoyar esta afirmación se ha tomado como ejemplo constante la expansión de la red vial de Brasil por toda la selva amazónica. Sin embargo la evidencia empírica de este efecto es escasa porque los datos sobre los cambios en la superficie boscosa son muy deficientes. Además, la mayoría de estudios se han realizado a una escala nacional y no regional lo que permitiría obtener una mejor estimación de las tasas de deforestación a nivel local.

La construcción de vías ha sido tradicionalmente una de las herramientas más importantes para el desarrollo rural y, por otra parte, se cree que favorecen la calidad de vida de la

población (Lipton & Ravallion, 1995). Chomitz & Gray (1996) afirman que el impacto de las carreteras es muy sensible a la calidad del suelo y a las regulaciones de tenencia de la tierra. Algunas nuevas vías favorecerán la tala del bosque para dar paso a cultivos comerciales, mientras que otras, estimularán la expansión de la agricultura migratoria, por lo tanto los autores hacen énfasis en la necesidad del uso de mapas de suelos para evaluar los patrones espaciales de amenaza, ya que esto proporcionaría información útil para el diseño de programas integrados de desarrollo y conservación.

Las evaluaciones ambientales como medio para mitigar los impactos negativos de proyectos viales han contribuido a la protección de los espacios naturales. Sin embargo no han logrado detener la pérdida de biodiversidad. En muchos países en desarrollo, las evaluaciones de impacto ambiental (EIA) se centran exclusivamente en la ruta que atraviesa la vía ignorando los impactos en áreas cercanas a las vía. Por ejemplo en Brasil, la vía Belén-Brasilia, de 2.000 km de largo finalizada a principios de los años 70, ha evolucionado hasta convertirse en una franja de 400 km de ancho en deforestación (Laurance et al., 2009). Por lo tanto es muy importante proyectar la condición a largo plazo de los bosques mediante una serie de predictores biofísicos (Soares-Filho et al., 2006), para determinar los patrones espaciales de la pérdida de bosque por la construcción de nuevas vías.

Muchos métodos se han desarrollado para llevar a cabo la planificación en la conservación de forma sistemática (Wilson et al., 2005b). Sin embargo, el riesgo de sistemas naturales con respecto a un fenómeno de origen antrópico como la deforestación, podrían afectar el desarrollo de este tipo de métodos al impedir el logro de los objetivos de conservación. Por lo tanto es importante realizar análisis comparativos de los principales procesos de cambio en la cobertura de la tierra y métodos avanzados para monitorear y modelar los cambios de ocupación del suelo a escala regional.

Varios métodos genéricos y herramientas podrían aplicarse para evaluar los procesos de deforestación. Uno de esos métodos es el uso de indicadores. Un indicador proporciona información sobre un asunto de mayor importancia o hace perceptible una tendencia o fenómeno que no lo era. Estos indicadores permiten describir el riesgo de los ecosistemas ante un proceso de amenaza como la deforestación, por lo que deben ser elegidos de una forma adecuada para resumir y medir la información pertinente sobre el riesgo (Locatelli et al., 2008). Los indicadores son por tanto útiles para explicar la complejidad de un fenómeno en términos simples y realizar comparaciones tanto temporales como espaciales.

Otras herramientas más cuantitativas son los modelos que utilizan aproximaciones estadísticas, los cuales contribuyen a entender cómo ciertas variables interactúan y controlan los patrones de deforestación (Geist & Lambin, 2001). Por ejemplo en el estudio realizado por Cropper et al. (2001) se estimó un modelo probit para explicar la deforestación y la ubicación de las áreas protegidas en el norte de Tailandia. Como variables explicadoras de la deforestación se incluyeron en el modelo variables fisiográficas (elevación, pendiente y tipo de suelo) y variables *dummy* para representar áreas correspondientes a parques nacionales y santuarios de fauna. El estudio permitió identificar aquellos sitios en el norte de Tailandia, en los que la construcción de vías influyó significativamente la deforestación.

En Chile se usaron modelos de regresión logística, para evaluar la vulnerabilidad de los bosques a una amenaza específica (Wilson et al., 2005a). Amenazas como las originadas por la fragmentación de los bosques, la conversión de bosques para uso agrícola, el aprovechamiento forestal y el fuego. Este estudio resalta la importancia de la utilización de técnicas estadísticas para evaluar la susceptibilidad de los bosques existentes a un proceso de amenaza específica y de cómo las diferentes técnicas pueden complementarse entre sí .

En la Amazonia colombiana se realizó un análisis espacial para evaluar la dinámica de la deforestación con un método de zonificación de la cobertura boscosa (Etter et al., 2006b). Esto permitió evaluar la velocidad de colonización, comparar patrones de deforestación y regeneración de bosques, así como la identificación de áreas a nivel regional que experimentaron un acelerado proceso de deforestación y regeneración. En este estudio se hace énfasis en la importancia de identificar lugares que son vulnerables a la deforestación, dada la dificultad en el diseño y puesta en marcha de medidas de conservación en un escenario de rápida deforestación.

Los métodos estadísticos espaciales se han desarrollado para facilitar el monitoreo de patrones geográficos y han sido útiles para predecir cambios en el uso del suelo por el mejoramiento y construcción de vías. Por ejemplo, en la región central de México se realizó un estudio para simular el impacto de vías sobre deforestación (Nelson & Hellerstein, 1997). Los resultados del estudio indican que el acceso a vías influyó significativamente el cambio en el uso de la tierra.

En la provincia del Darién Panamá, Nelson et al. (1999) usaron un modelo logit para predecir cambios en el uso de la tierra como resultado del mejoramiento de la vía Panamericana. Los resultados indican que la repavimentación de la vía tendría poco efecto sobre la deforestación. Estos resultados luego se confirmaron en un estudio posterior en el cual se estimaron tres modelos: logit multinomial, logit anidado y logit con parámetros aleatorios (Nelson et al., 2004). El estudio concluyó que el mejoramiento de la vía no produciría cambios significativos en los patrones de deforestación independientemente del tipo modelo empleado.

Además de las técnicas y modelos estadísticos mencionados anteriormente, existen métodos computacionales como las redes neuronales artificiales (RNAs) que permiten realizar predicciones ante la presencia de relaciones no lineales. Por lo tanto son útiles para modelar fenómenos naturales, cuando la relación entre variables no es conocida



(Razi & Athappilly, 2005). Por ejemplo en el estudio de la deforestación, al tener en cuenta factores ambientales, socioeconómicos y culturales, la relación entre variables puede ser muy compleja, ya que la deforestación no es el resultado de la suma de cada factor en una forma independiente, sino una combinación de ellos.

Las Redes Neuronales Artificiales (RNAs) se han utilizado ampliamente en muchos campos de investigación, principalmente en la computación y neurofisiología. Recientemente se han usado en áreas como la biología y las ciencias ambientales, para predecir la distribución de las especies, la abundancia o la diversidad en función de variables ambientales (Lek-Ang et al., 1999; Manel et al., 1999) y para caracterizar los ecosistemas a partir de la interpretación de imágenes satelitales (Jensen et al., 1999; Paruelo & Tomasel, 1997). Este modelo fue utilizado por Mas et al. (2004) para predecir cambios en el uso del suelo en el estado de Campeche en el sur este de México. El estudio resalta las ventajas de las RNAs frente a los métodos estadísticos convencionales, como la regresión logística, los cuales se han utilizado en varios trabajos de deforestación en la región tropical.

Los métodos estadísticos como el análisis de regresión y el análisis multivariado se han aplicado a una amplia gama de decisiones en muchas disciplinas. Estos modelos son atractivos para los tomadores de decisiones, debido a su estabilidad metodológica y la larga historia de aplicación. Sin embargo para este tipo de modelos se deben realizar hipótesis a priori acerca de la relación entre las variables, que podrían no existir. La principal ventaja en el uso de las RNAs para la predicción, es que este tipo de suposiciones no es necesario, por lo tanto, pueden ser un instrumento más apropiado cuando se pone énfasis en la propia predicción y no en las relaciones subyacentes entre variables independientes y dependientes (Brey et al., 1996).

Varios investigadores han comparado los modelos de redes neuronales con modelos logit utilizando diferentes conjuntos de datos, encontrando un rendimiento superior para las redes neuronales. Por ejemplo Yilmaz (2009) utilizó RNAs, un modelo de relación de

frecuencias y un modelo logit para estimar la susceptibilidad a los deslizamientos en la provincia de Tokat al norte de Turquía. Aunque los resultados del estudio fueron similares para los tres modelos, el análisis del área bajo la curva ROC, indicó un mejor rendimiento en la predicción con RNAs comparado con los otros dos métodos.

Goss & Vozikis (2002) compararon RNAs con métodos de regresión logística para predecir las tasas de mortalidad de pacientes en cuidados intensivos y concluyeron que la precisión de la predicción a partir del modelo de RNAs fue mejor que la del modelo de regresión logística. Hruschka (1993) realizó una comparación de técnicas econométricas con modelos de RNAs aplicados a funciones de respuestas del mercado. El autor indica que el modelo de RNAs utilizando el algoritmo de retropropagación del error, podría conducir a un mejor ajuste comparado con el modelo econométrico. Sin embargo, resalta la necesidad de realizar más estudios para establecer conclusiones generales con respecto a las fortalezas y debilidades de las redes neuronales. De igual forma Warner & Misra (1996) quienes compararon el desempeño de las redes neuronales con análisis de regresión, discuten las situaciones en las que sería ventajoso el uso de modelos de RNAs en lugar de un modelo de regresión.

Las RNAs y los modelos estadísticos se han desarrollado prácticamente de forma independiente. Sin embargo, algunos investigadores que han examinado modelos de RNAs desde una perspectiva estadística indican que considerar principios estadísticos en el proceso de construcción de modelos RNAs puede mejorar el rendimiento del mismo. Incluso la combinación de los dos métodos mejoraría las estimaciones en la predicción. Por ejemplo Zhang (2003) desarrolló un modelo híbrido para la predicción de series de tiempo mediante la combinación de ecuaciones lineales y no lineales. El modelo híbrido propuesto, utiliza primero un modelo autorregresivo de media móvil (ARMA) para modelar la parte lineal del sistema y luego una red neuronal unidireccional, entrenada con el error encontrado por la ecuación de ARMA. La ecuación ARMA se utiliza para modelar los aspectos lineales, dejando los elementos no lineales para la red neuronal. Los modelos

se sumaron para generar la predicción completa y los resultados mostraron que el modelo híbrido superó a ambos modelos de forma independiente.

Las redes neuronales también se han usado con modelos de autómatas celulares (AC) para predecir cambios en el uso del suelo, que se caracterizan por la interacción de los patrones de paisajes vecinos (Pan et al., 2010). El modelo de RNAs-AC se usó para la simulación de múltiples cambios de uso del suelo en la ciudad de Dongguan, en el sur de China (Li & Yeh, 2002). El estudio muestra como el método propuesto puede superar algunas de las deficiencias de los modelos AC utilizados actualmente en la simulación de sistemas urbanos complejos y múltiples cambios de uso del suelo, por reducir de forma significativa el trabajo de definir el valor de los parámetros, las reglas de transición y la estructura del modelo.

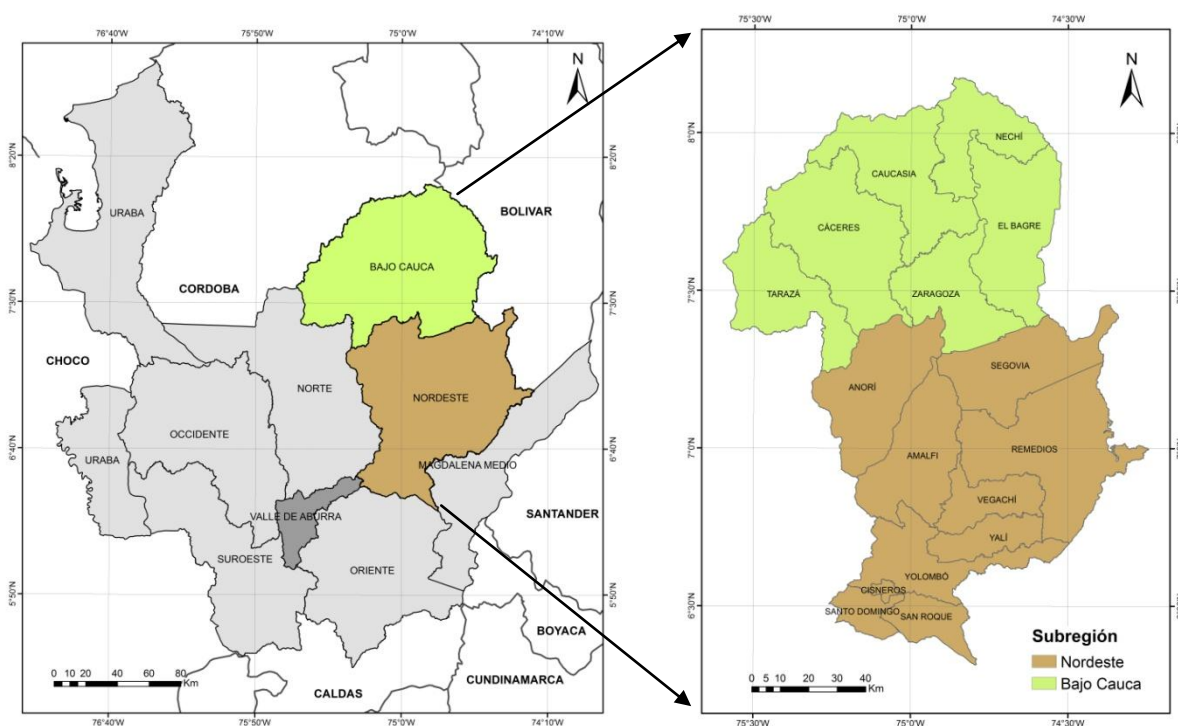
En este contexto desarrollar un modelo que simule la deforestación, es muy importante para proyectar diferentes escenarios de cambio en el uso del suelo. Además, su elaboración lleva a un mejor conocimiento del fenómeno y permite estimar los patrones y las tasas de deforestación en función de diversos parámetros. Modelar el proceso de deforestación se basa en la hipótesis de que el uso del suelo y los cambios de uso están inducidos por individuos o grupos de individuos que toman decisiones racionales condicionadas por factores medioambientales y socioeconómicos. Por estas razones, la acción de estos factores no es aleatoria y se puede modelar (Meyer & Turner, 1994).

Los modelos cuantitativos de cambio de uso del suelo podrían aplicarse en una variedad de propósitos de planificación ambiental, que ayudarían a la toma de decisiones y a mejorar las actividades con el fin de mitigar los impactos de nuevos proyectos viales. Además son útiles para generar escenarios futuros de cambios en el uso del suelo, para predecir la localización de la deforestación y para apoyar el diseño de políticas públicas orientadas a la conservación (Boyd, 1996).

## 6. METODOLOGÍA

### 6.1 Área de Estudio

El área de estudio corresponde a las regiones del Bajo Cauca y Nordeste del departamento de Antioquia (figura1) las cuales comprenden 6 y 10 municipios, con un área total de 8.585 km<sup>2</sup> y 8.645km<sup>2</sup> respectivamente. Estas subregiones aún conservan grandes áreas en bosque con gran diversidad de recursos naturales y están dedicadas principalmente a la ganadería y a la minería con producción agrícola marginal (Gobernación de Antioquia et al., 2007). Su ubicación geográfica ha conducido a la planeación de nuevos proyectos de infraestructura vial con el fin de mejorar la conectividad del departamento con el norte del país, lo cual pone en riesgo la conservación de los bosques que aún se encuentran en esta zona, por lo tanto el área de estudio constituye una región importante para simular y evaluar el posible efecto de la construcción de nuevas vías sobre la deforestación.



**Figura 1.** Localización del área de estudio correspondiente a las subregiones del Nordeste y Bajo Cauca del departamento de Antioquia.

## **6.2 Datos y Variables**

Luego de revisar la literatura sobre los principales determinantes de la deforestación en la región tropical, se realizó una rigurosa búsqueda de información espacialmente explícita. La información colectada se utilizó para la construcción de las variables explicadoras de la deforestación, las cuales representan características biofísicas y de accesibilidad que prevalecían lo más cerca posible al año 1980, año que corresponde al inicio del periodo de análisis de la presente investigación.

### **6.2.1 Coberturas Terrestres de los años 1980 y 2000.**

La información de coberturas terrestres para el año 1980 se obtuvo luego de digitalizar en el Laboratorio de Bosques y Cambio Climático de la Universidad Nacional de Colombia, Sede Medellín, más de 500 planchas temáticas (Orrego 2009) en formato análogo, a escala 1:25.000, proporcionadas por la Secretaría de Agricultura de Antioquia.

La información de coberturas terrestres para el año 2000 fue el resultado de un Convenio Interadministrativo entre la Secretaría de Agricultura, el Departamento Administrativo de Planeación y el Instituto Geográfico Agustín Codazzi. Aproximadamente 8000 fotografías aéreas y nueve imágenes de satélite, 8 Landsat 7 ETM+ y una SPOT 5, se usaron para realizar un levantamiento semidetallado de coberturas terrestres en Antioquia, a escala 1:25.000. El estudio de coberturas para el año 2000 constituyó una adaptación de la metodología CORINE Land Cover (Coordination of Information of the Environment).

La información de coberturas para ambos periodos se agrupó en siete tipos de coberturas terrestres: bosque denso (BD), bosque mixto (BM), agricultura (A), pastos (P), plantaciones forestales (PF), vegetación arbustiva (Ar) y otras coberturas (OC) (Orrego, 2009). El bosque denso corresponde a coberturas forestales dominadas por árboles, con un dosel continuo, la cobertura de bosque mixto incluyó áreas de mosaicos en las que los bosques se encuentran mezclados con vegetación leñosa de porte alto o bajo. En la categoría de agricultura se incluyen tanto cultivos temporales como permanentes. Finalmente, en otras

coberturas se incluyen áreas urbanas, cuerpos de agua, suelos erosionados, áreas mineras, áreas rocosas, pantanos, áreas de recreación y áreas de explotación de petróleo.

**6.2.2 Vías:** se obtuvo de la Secretaría de Infraestructura de Antioquia un archivo digital en formato vector de las vías primarias, secundarias y terciarias a escala 1:25.000.

**6.2.3 Hidrografía:** se obtuvo de la Secretaría de Agricultura de Antioquia un archivo de la red de drenaje para el departamento en formato vector a escala 1:25.000.

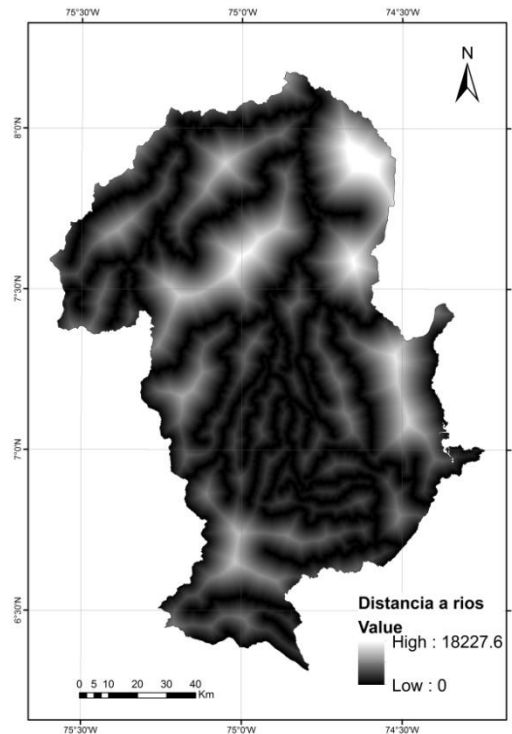
**6.2.4 Modelo de elevación digital del terreno:** se usó el modelo de elevación digital ASTER con una resolución aproximada de 30m (METI & NASA, 2009).

**6.2.5 Suelos:** se obtuvo el mapa digital de suelos del año 1979, a escala 1:25.000 proporcionado por la Secretaría de Agricultura de Antioquia. El mapa contiene las categorías de clases agrológicas definidas por el Departamento de Agricultura de los Estados Unidos (USDA), datos de fertilidad, drenaje y pedregosidad. La clasificación de clases agrológicas se basa en factores como la fertilidad del suelo, la profundidad, la textura, la pendiente, la salinidad y la toxicidad química.

### **6.3 Creación de variables explicadoras de la deforestación**

#### **6.3.1 Distancia euclidiana a ríos**

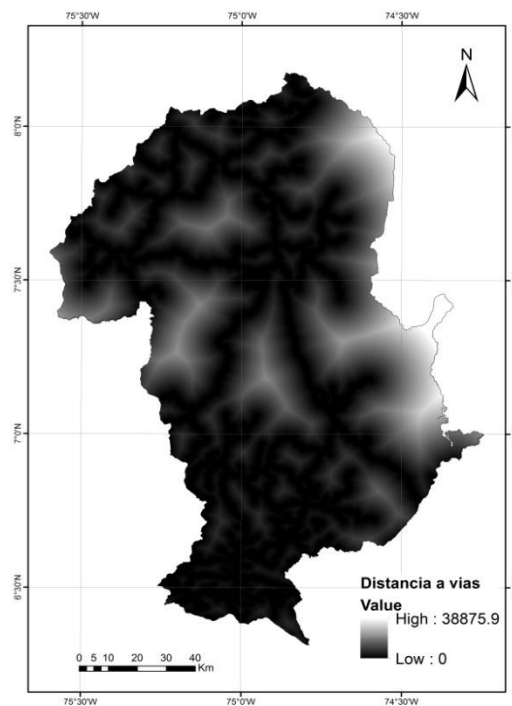
Con la red de drenaje a escala 1:25.000 se generó un archivo en formato raster con la distancia euclidiana a los principales ríos, es decir ríos navegables que facilitan el transporte de productos agrícolas y la conectividad entre los municipios del área de estudio (figura 2).



**Figura 2.** Distancia euclidiana a ríos. Tamaño de pixel 30m

### 6.3.2 Distancia euclidiana a vías

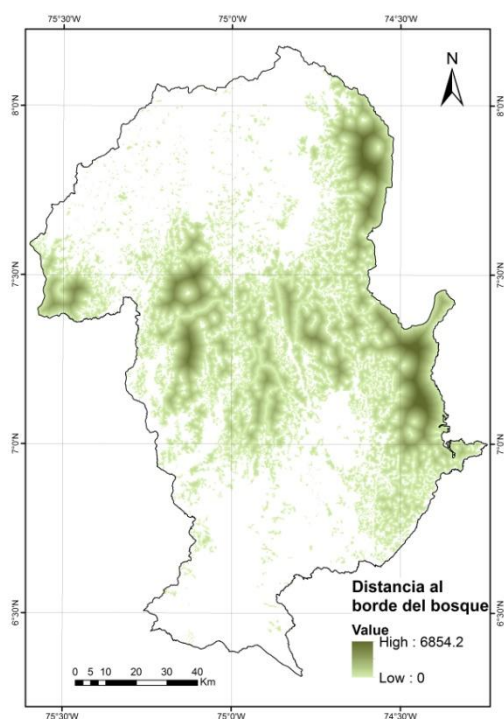
Con la información de la red vial primaria, secundaria y terciaria se generó un archivo en formato raster con la distancia euclidiana a las vías (figura 3).



**Figura 3.** Distancia euclidiana a vías. Tamaño de pixel 30m

### 6.3.3 Distancia al borde del bosque

Se utilizó la información en formato vector de la cobertura boscosa (bosque denso y bosque mixto) y no boscosa para el año 1980 y se calculó la distancia euclidiana desde la cobertura en bosque hacia la cobertura sin bosque, obteniendo finalmente la distancia al borde del bosque. Estudios han mostrado como la deforestación comienza desde el borde del bosque o cerca de sitios previamente deforestados, por lo tanto se considera una variable importante para modelar los patrones de deforestación (Thies et al., 2012).



**Figura 4.** Distancia al borde del bosque/No bosque. Tamaño de pixel 30m

### 6.3.4 Distancia de menor costo a las cabeceras municipales

La accesibilidad a mercados podría medirse de forma más adecuada con los tiempos de viaje que con las distancias lineales. Sin embargo, la información de tiempos de viaje no está generalmente disponible. Por esto los estudios de deforestación tropical usan distancias ponderadas por la impedancia (Chomitz & Gray, 1996; Nelson & Hellerstein, 1997; Nelson et al., 2004). La impedancia o costo, representa la dificultad de moverse planimétricamente a través de cada celda o pixel, por lo tanto el cálculo de esta variable



equivale a la identificación de la distancia de menor costo desde un pixel a cada uno de los mercados, caracterizados espacialmente por las cabeceras municipales existentes en 1980.

En la presente investigación se asignaron valores de impedancia a las coberturas, vías, ríos, áreas protegidas y pendiente, teniendo en cuenta las estimaciones del costo de transporte de productos agrícolas realizadas en el estudio de Nelson et al. (1999), quienes utilizaron técnicas de análisis espacial para predecir cambios en el uso del suelo que podrían ocurrir luego de la repavimentación de la carretera Panamericana ubicada al extremo Sureste de Panamá.

- **Asignación del valor de impedancia por tipo de cobertura:** A las áreas forestales, que incluyen todos los tipos de bosques y aquellas cubiertas con vegetación arbórea de porte bajo, se les asignó un valor de impedancia 5. A las áreas con uso agropecuario, plantaciones forestales y áreas urbanas, se les asignó un valor de impedancia 3. Para la asignación de valores de impedancia se asume que las áreas con bosques presentan mayores dificultades para el transporte de productos agrícolas, por tanto, estas áreas no se podrían cruzar relativamente fácil si se comparan con áreas que tienen otro tipo de cobertura como pastos o zonas urbanas.
- **Asignación del valor de impedancia por tipo de vía:** Un valor de impedancia 1 se asignó a las vías pavimentadas las cuales constituyen nodos importantes en la red de transporte terrestre de la región. Un valor de impedancia 3 a las vías sin pavimentar, un valor de impedancia 4 a caminos de herradura que solo permiten el tránsito de bestias de carga y un valor de impedancia 5 a sitios sin vías para representar la mayor dificultad en el transporte de productos agropecuarios.
- **Asignación del valor de impedancia a los ríos:** los ríos podrían tener un valor de impedancia similar al de una vía (Nelson et al., 1999) ya que representan una

importante fuente de transporte y proporcionan fácil acceso a los principales centros urbanos. Por tanto a los ríos se les asignó un valor de impedancia 1 y al resto del área un valor de 5.

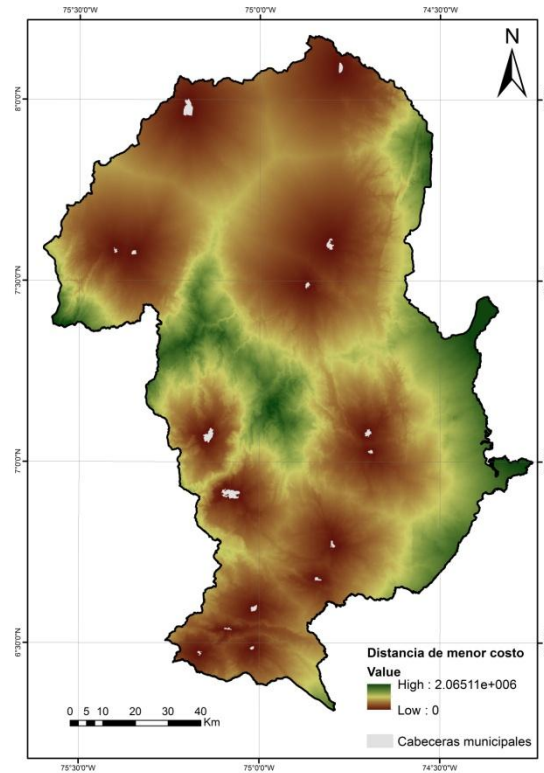
- **Asignación de impedancia a las áreas protegidas:** aunque es posible trasportar productos agrícolas a través de un área protegida, este tipo de áreas representan una barrera natural para el transporte por encontrarse bajo protección legal. Por lo tanto al área protegida presente en la zona de estudio, la cual corresponde a la Reserva de Recursos Naturales Bajo Cauca-Nechí, se le asignó un valor de impedancia de 5 y al resto del área de estudio un valor de 1.

Finalmente para obtener la superficie de costo, los valores de impedancia asignados se sumaron y luego se multiplicaron por una función de la pendiente sugerida por Nelson et al. (1999).

$$\left(1 + \frac{P_e^2}{50}\right) * (Cc + Cv + Cr + Ca) \quad (1)$$

Donde  $P_e$  es el grid de pendiente,  $Cc$  el costo por tipo de cobertura,  $Cv$  el costo por tipo de vía,  $Cr$  el costo asignado a los ríos y  $Ca$  al costo asignado al área protegida.

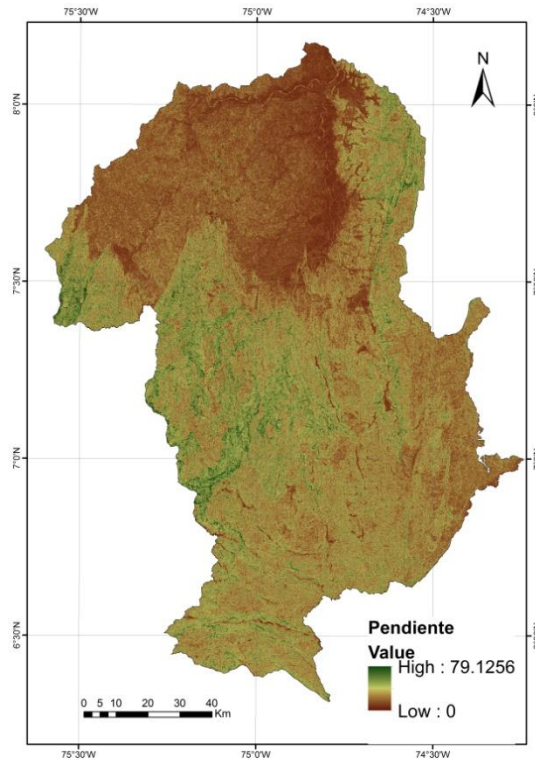
Aunque la fórmula es un poco arbitraria, captura la relación positiva entre el aumento en la pendiente y el mayor costo de accesibilidad a mercados.



**Figura 5.** Distancia de menor costo a las cabeceras municipales. Tamaño de pixel 30m

### 6.3.5 Pendiente

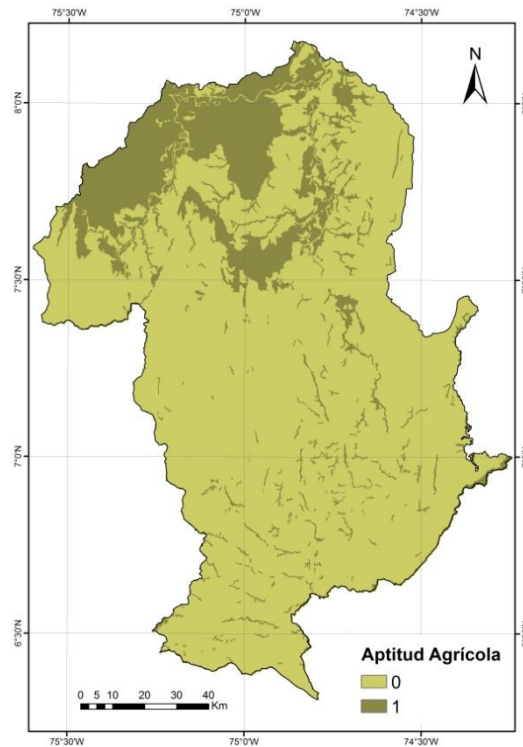
Se generó un mapa en formato raster de la pendiente en grados (figura 6) a partir del modelo de elevación digital del terreno (DEM). Se considera que tierras planas son más aptas para el uso agropecuario que aquellas ubicadas en topografías montañosas, por poseer suelos más fértiles y facilitar el uso de maquinaria agrícola.



**Figura 6.** Pendiente (grados). Tamaño de pixel 30m

### 6.3.6 Aptitud Agrícola

Con el mapa digital de suelos para el año 1979 se creó una variable *dummy* en formato raster con la información de clases agrológicas (figura 7). Se asignó un valor de 1 para los píxeles en las clases II y III, es decir, para aquellas zonas apropiadas para un uso agrícola y cero para las otras clases (IV,V,VI,VII,VIII). Se asume que áreas con mayor potencial agrícola son más susceptibles a deforestación.



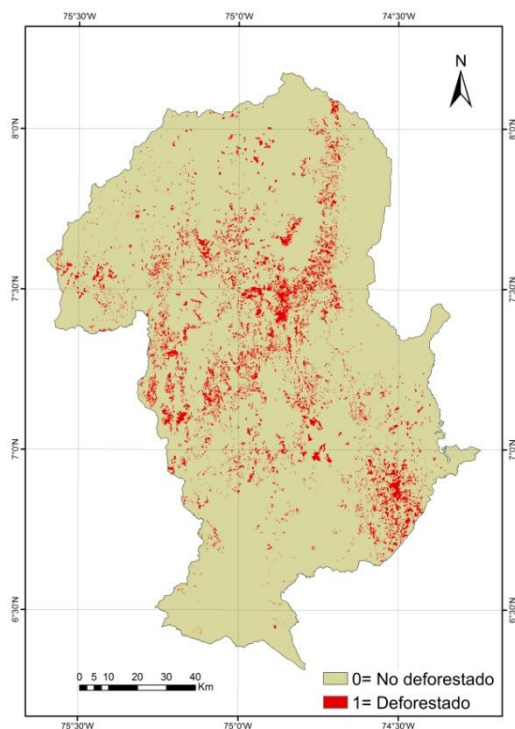
**Figura 7.** Aptitud Agrícola. Tamaño de píxel 30m

Aunque la deforestación puede ser explicada por una gran cantidad de variables, las causas subyacentes de la deforestación en la zona de estudio (pobreza, pobre gestión forestal, la inseguridad de tenencia de tierras etc.) las cuales constituyen las condiciones iniciales en las relaciones humano-ambientales, son más difíciles de identificar que las causas próximas debido a la disponibilidad de datos. Por lo tanto se utilizaron variables biofísicas y de accesibilidad que permiten una mejor representación espacial. Sanabria (2009) y Müller & Mburu (2009) también reportaron esta limitación y utilizaron solo este tipo de variables para modelar la deforestación con RNAs y a partir de métodos estadísticos como la regresión logística.

#### **6.4 Creación de la variable dependiente**

Por deforestación se entiende el cambio de bosque a no bosque. Esto corresponde al área en bosque en 1980 que veinte años después se convirtió a usos alternativos, como

agricultura, pastos, vegetación arbustiva y otras coberturas. Por lo tanto para generar la variable deforestación se realizó una intersección entre los archivos en formato vector de la cobertura boscosa presente en 1980 (bosque denso y bosque mixto) con el archivo de cobertura no boscosa para el año 2000. Luego el archivo generado se convirtió a raster y se reclasificó asignando valores de cero para aquellos sitios que no se deforestaron, y valores de uno en caso contrario (figura 8).



**Figura 8.**Variable dependiente. Tamaño de pixel 30m

La construcción de las variables se realizó en el software ArcGis 10.1 (ESRI, 2010), con un tamaño de pixel de 30 m ya que de acuerdo con Orrego (2009) es el más adecuado para la representación en formato raster de la variable deforestación y de aquellas que explican su ocurrencia.

## 6.5 Tasa de deforestación

Con la información del área cubierta con bosque tanto en 1980 como en el 2000, se calculó la tasa bruta de deforestación para toda el área de estudio, empleando la fórmula propuesta por Puyravaud (2003).

$$tasa_i = \frac{\ln A_2 - \ln A_1}{t_2 - t_1} \quad (2)$$

Donde  $A_1$  y  $A_2$  representan la cobertura de bosque en el tiempo  $t_1$  y  $t_2$  respectivamente. La tasa se denomina bruta por no considerar el área que experimentó la recuperación de su cobertura forestal.

## **6.6 Modelación de la deforestación**

Para evaluar la importancia relativa de las variables explicadoras sobre la deforestación se estimó un modelo de regresión logística y con el conjunto de variables que presentaron mejor predicción se realizó un modelo con RNAs, ya que de acuerdo con la literatura éste método genera mejores predicciones porque tiene la capacidad de capturar relaciones no lineales (Rojas, 1996a; Mas et al., 2004).

### **6.6.1 Modelo de Regresión Logística**

La regresión logística es un modelo de análisis multivariado, útil para predecir la presencia o ausencia de una característica o resultado, basado en los valores de un conjunto de variables predictoras. La ventaja de la regresión logística es la posibilidad de incluir en el modelo variables continuas o discretas y cualquier combinación de ellas sin la necesidad de presentar una distribución normal (Paruelo & Tomasel, 1997).

Para estimar el modelo se seleccionó una muestra aleatoria de 700.000 puntos (10% de los datos) donde cada punto representa un pixel de 30 metros y una segunda muestra de 20.000 puntos en la que se evitó la inclusión de observaciones espacialmente contiguas, para controlar el posible efecto de la autocorrelación espacial en la estimación de los parámetros del modelo.

Las variables explicadoras descritas anteriormente se incorporaron al modelo como variables independientes y la deforestación observada durante el periodo 1980-2000, como variable dependiente de tipo binaria o dicotómica.

Dado que la multicolinealidad podría estar presente en modelos estadísticos, se estimó una matriz de correlación para cada muestra con el fin de identificar las variables a incluir en el modelo. Una correlación de por lo menos 0,6 se consideró como un valor indicativo de alta correlación entre pares de variables.

Dada una función lineal de la forma.

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (3)$$

Donde  $Y$  es la variable dependiente y representa la presencia (1) o ausencia (0) de deforestación,  $b_0$  el intercepto,  $b_1 \dots b_n$  los coeficientes de regresión y  $x_1 \dots x_n$  las variables explicadoras.

La probabilidad de deforestación se calculó a partir de la ecuación 4:

$$x = \{x_1, x_2, \dots, x_n\} \quad P_i = \frac{1}{1 + \exp(-\beta_0 - \beta_1X_1 - \beta_2X_2 - \dots - \beta_nX_n)} \quad (4)$$

Donde  $x$  representa el conjunto de variables explicadoras y  $P_i$  la función de distribución acumulada de una variable que se distribuye acorde a una función logística. Los modelos se estimaron en el software RStudio (R Development Core Team, 2008).

## 6.6.2 Modelo de Red Neuronal

### 6.6.2.1 Marco Conceptual de Redes Neuronales Artificiales

Las Redes Neuronales Artificiales, RNAs (en inglés, *Artificial Neural Networks*) están inspiradas en las redes neuronales biológicas del cerebro humano. No obstante conviene aclarar que esta inspiración no supone que las RNAs lleguen a emular al cerebro, simplemente se componen de elementos que se comportan de forma similar a la neurona biológica en sus funciones más básicas, como por ejemplo aprenden de la experiencia,



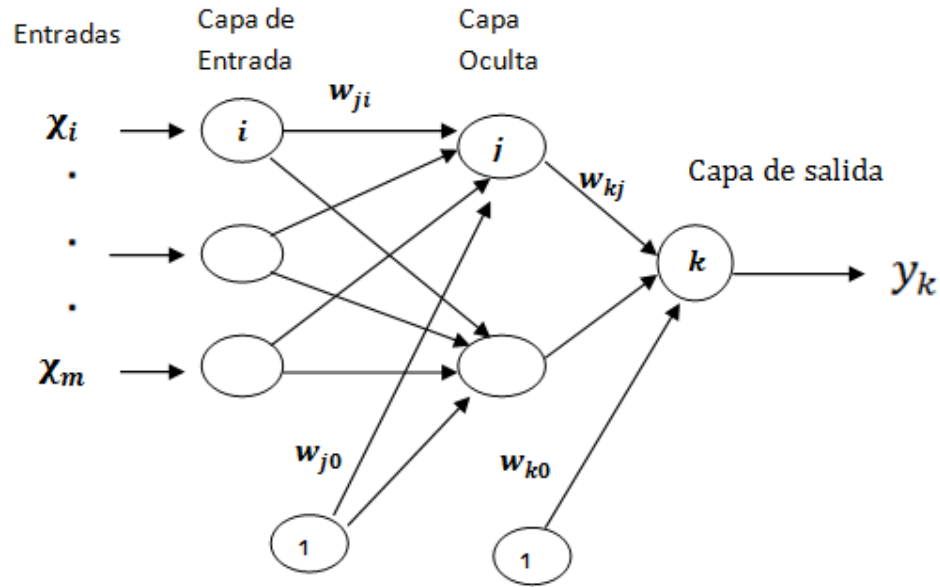
generalizan de ejemplos previos a ejemplos nuevos y abstraen las características principales de una serie de datos (Bishop, 1995).

El primer modelo de red neuronal fue propuesto por McCulloch & Pitts (1943) en términos de un modelo computacional de actividad nerviosa. Este modelo sirvió de base para los modelos posteriores que han sido aplicados en diferentes campos de investigación. A menudo las RNAs se aplican en la medicina y biología molecular. A comienzos de 1990 se reportaron los primeros estudios en el área de ecología y ciencias ambientales (Lek & Guégan, 1999).

Colasanti (1991) encontró similitudes entre las RNAs y los ecosistemas y recomienda la utilización de esta herramienta en modelos ecológicos. Comrie (1997) estudió el potencial de las redes neuronales para predecir la contaminación por ozono y Hilbert & Ostendorf (2001) utilizaron RNAs para modelar los impactos del cambio climático sobre la distribución de la vegetación. En este trabajo se describe un conjunto de métodos aplicados en los trópicos húmedos del norte de Queensland en los cuales se utiliza la distribución actual de los bosques, variables biofísicas y las RNAs para indagar sobre los patrones de vegetación en el clima del pasado, presente y futuro. Un resumen del uso de RNAs en diferentes áreas de investigación se encuentra en (Razi & Athappilly, 2005).

La investigación sobre las RNAs ha llevado al desarrollo de distintos tipos de redes capaces de resolver diferentes problemas como memoria auto-asociativa, generalización, optimización y predicción. Por lo tanto, la elección del tipo de red depende de la naturaleza del problema a resolver. En la presente investigación se usó una red neuronal unidireccional también llamada perceptrón multicapa (figura 9) con el algoritmo de retropropagación por que es el tipo de red más empleado para realizar tareas de predicción y generalización. Este tipo de red actúa como un aproximador universal de funciones y tiene la capacidad de dar salidas satisfactorias a entradas que la red no ha visto en su fase de entrenamiento (Rojas, 1996b).

El desarrollo operativo de ésta red presenta dos etapas, una de funcionamiento y la otra de entrenamiento. Durante la etapa de funcionamiento se presenta ante la red un patrón de entrada y éste se transmite a través de las sucesivas capas de neuronas hasta obtener la salida. El número de neuronas de la capa de entrada está determinado por el número de variables explicadoras, el número de neuronas de la capa oculta determina la capacidad de aprendizaje de la red neuronal y finalmente la capa de salida posee una neurona la cual corresponde a la variable dependiente.



**Figura 9.** Esquema de un Perceptrón Multicapa. En este ejemplo  $X_i \dots X_m$  corresponden a las variables explicadoras,  $w_{ji}$  el peso de conexión entre la neurona de entrada  $i$  y la neurona oculta  $j$ ,  $w_{kj}$  el peso de conexión entre la neurona oculta  $j$  y la neurona de salida  $K$  y El peso  $w_{j0}$  se considera como un peso asociado a una neurona ficticia con un valor de salida igual a 1.

De acuerdo con la grafica anterior la entrada total o neta que recibe una neurona oculta  $j$ ,  $v_j$  para la iteración  $n$  está dada por (Haykin ,1999).

$$v_j(n) = \sum_{i=0}^m w_{ji}(n)x_i(n) \quad (5)$$

El sesgo aplicado a las neuronas ocultas y de salida está representado por los pesos  $w_{j0}$  y  $w_{k0}$ . Este término que también se denomina “bias” mejora las propiedades de convergencia de la red y actúa de manera similar al término constante en los modelos de regresión.

El valor de salida de la neurona oculta  $j$ ,  $y_j(n)$ , se obtiene aplicando una función sobre su entrada neta:

$$y_j(n) = \varphi_j(v_j(n)) \quad (6)$$

Donde  $\varphi_j(\cdot)$  por conveniencia matemática corresponde a la función logística definida como:

$$y = \frac{1}{1+e^{-x}} \quad (7)$$

De igual forma, la entrada neta que recibe una neurona de salida  $k$ ,  $v_k$  es:

$$v_k(n) = \sum_{i=0}^H w_{ki}(n)y_i(n) \quad (8)$$

Donde  $H$  corresponde al número de neuronas en la capa oculta.

Por último, el valor de salida de la neurona de salida  $k$ ,  $y_k$  esta dado por:

$$y_k(n) = \varphi_k(v_k(n)) \quad (9)$$

Luego de la etapa de funcionamiento, inicia la etapa de entrenamiento, en la cual el algoritmo de retropropagación modifica los pesos de la red de forma iterativa con el objetivo de hacer mínimo la función del error, la cual es una medida del comportamiento de la red neuronal, la función está representada por la siguiente ecuación (Haykin ,1999).

$$E_v = \frac{1}{N} \sum_{n=1}^N E(n) \quad (10)$$

Donde  $N$  representa el número total de patrones presentados a la red.

$E(n)$  está dado por:

$$E(n) = \frac{1}{2} \sum_{j \in C} e_k^2(n) \quad (11)$$

Donde  $C$  representa el número de neuronas en la capa de salida

$$e_k(n) = d_k(n) - y_k(n), \quad (12)$$

El algoritmo de retropropagación usa la técnica conocida como gradiente decreciente. Dado que el gradiente toma la dirección que determina el incremento más rápido en el error, la dirección opuesta, es decir, la dirección negativa, determina la dirección en la cual el error disminuye más rápidamente. Por tanto, el error puede reducirse ajustando cada peso en la dirección en la que decrece el error (Rojas, 1996b).

Los pesos se modifican de forma iterativa al aplicar la regla de la cadena a la expresión del gradiente y añadir una tasa de aprendizaje  $\eta$ , la cual controla el tamaño del cambio de los pesos en cada iteración, así el cambio en los pesos de una neurona de salida está dado por (Haykin, 1999).

$$\Delta w_{kj} = \eta * \delta_k(n) * y_k(n) \quad (13)$$

Donde: 
$$\delta_k(n) = e_k(n) \varphi'_k(v_k(n)) \quad (14)$$

El gradiente local  $\delta_k(n)$  para la neurona de salida  $k$  es igual al producto del correspondiente error  $e_k(n)$  para esa neurona y la derivada  $\varphi'_k(v_k(n))$  de la función de activación asociada.

En una neurona oculta el cambio de los pesos está dado por:

$$\Delta w_{ji} = \eta * \delta_j(n) * y_j(n) \quad (15)$$

Donde: 
$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (16)$$

De acuerdo con lo anterior, el error o valor delta asociado a una neurona oculta  $j$ , está determinado por la suma de los errores que se cometen en las  $k$  neuronas de salida, que reciben como entrada la salida de esa neurona oculta  $j$ . De ahí que el algoritmo también se denomine propagación del error hacia atrás.

Para acelerar el proceso de convergencia de los pesos, se recomienda añadir un factor momento,  $\alpha$  (Rumelhart et al., 1986) que tiene en cuenta la dirección del incremento tomada en la iteración anterior:

$$\Delta w_{kj}(n+1) = \eta * \delta_k(n) * y_j(n) + \alpha * \Delta w_{kj}(n) \quad (17)$$

$$\Delta w_{ji}(n+1) = \eta * \delta_j(n) * y_i(n) + \alpha * \Delta w_{ji}(n) \quad (18)$$

Las aplicaciones de aprendizaje supervisado usando el algoritmo de retropropagación se pueden dividir en dos categorías, aprendizaje por lotes (batch), en el cual la actualización de los pesos se realiza luego de presentar todos los datos o patrones a la red y el aprendizaje en línea (on line), en el cual los pesos se actualizan luego de presentar cada patrón a la red (Wilson & Martinez, 2003; Plagianakos et al., 2001; Rojas, 1996b). Este tipo de aprendizaje, también es llamado adaptativo ya que es útil para modelar un sistema variable en el tiempo (Plagianakos et al., 2001).

Varios estudios que han utilizado perceptrones múltiples con el algoritmo de retropropagación y una variedad de tipos de aprendizaje, apoyan la afirmación de que el modo de aprendizaje en línea (on line) es más rápido que el modo batch especialmente para grandes conjuntos de entrenamiento (Bishop, 1995; Bengio, 1996; Haykin, 1999).

Esto puede ser explicado por el hecho que el aprendizaje en línea puede utilizar una mayor tasa de aprendizaje por lo que la red se entrena más rápidamente, por el contrario, para grandes conjuntos de datos el entrenamiento en modo batch sería muy lento ya que sería necesario una tasa de entrenamiento mucho menor para mantener el aprendizaje estable (Wilson & Martinez, 2003).

Si bien el entrenamiento por lotes (batch) toma la dirección correcta del gradiente, el algoritmo en modo batch sólo puede dar un paso para cada época y cada paso es en línea recta, por lo tanto el gradiente se calcula solo en el punto de partida y no puede seguir las curvas de la superficie del error. Por el contrario, el entrenamiento on line utiliza el gradiente local de cada patrón para determinar qué dirección seguir en la superficie del error. Estos gradientes locales pueden ser erróneos y contradictorios entre sí, sin embargo al calcular el promedio, el algoritmo toma la dirección correcta del gradiente y permiten seguir las curvas de la superficie del error durante el curso de una iteración (Wilson & Martinez, 2003).

Varios investigadores han demostrado que para problemas complejos se debe utilizar un conjunto de entrenamiento grande para lograr una mayor precisión en la generalización (Haykin, 1999; Bengio, 1996). De acuerdo con lo anterior en la presente investigación se utilizó el método on line ya que tiene la capacidad de trabajar con grandes conjuntos de entrenamiento y datos redundantes, además la convergencia de la red es más rápida comparada con el método de aprendizaje en modo batch (Plagianakos et al., 2001).

#### **6.6.2.2 Entrenamiento de la red neuronal**

El entrenamiento se llevó a cabo en el software R Studio (R Development Core Team, 2013; RStudio, 2013) usando el paquete AMORE el cual hace uso del algoritmo de retropropagación (Castejón et al., 2010). Este paquete se desarrolló con base en la teoría presentada en el libro *“Neural Networks A Comprehensive Foundation”* por Haykin (1999).

AMORE es una herramienta destinada al entrenamiento y simulación de redes neuronales cuya diferencia con otras alternativas de extendido uso es la flexibilidad para adaptar la estrategia de aprendizaje, ya que permite obtener un control total de la red, acceder directamente a cada uno de los parámetros de la misma y personalizar sus diferentes funciones y características.

La red neuronal propuesta para iniciar la fase de entrenamiento presenta una estructura de perceptrón multicapa unidireccional (figura 9), con una capa de entrada, una capa

oculta y una capa de salida. Cada variable explicadora está asociada a una neurona en la capa de entrada y se asignó una neurona a la capa de salida que corresponde a la variable dependiente. Para esta última capa se usó una función de activación logística por tratarse de una variable categórica que toma valores de cero y uno.

De acuerdo con las características del paquete empleado, los pesos inician con valores aleatorios dentro del rango.

$$\pm \sqrt{\frac{3}{eo + os}} \quad (19)$$

Donde **eo** corresponde al número de conexiones entre la capa de entrada y la capa oculta y **os** al número de conexiones entre la capa oculta y la capa de salida.

El entrenamiento inicialmente se realizó con la misma muestra empleada en el modelo de regresión logística (700.000 puntos) en la cual el 86% corresponde a puntos que no se deforestaron y 14% a puntos deforestados, sin embargo con este tamaño de muestra el entrenamiento fue muy lento, por lo tanto se seleccionó una muestra balanceada de 250.000 puntos, aproximadamente el 3% de los datos. De esta muestra el 50% de los datos corresponde a puntos deforestados y el otro 50% a puntos que no se deforestaron. Este balance se realizó con el fin de mejorar el rendimiento de la red y lograr una mejor generalización, ya que para una muestra desbalanceada el aprendizaje de la red tiende a estar sesgado hacia la clase que presenta mayoría en el grupo de entrenamiento (Nguyen et al., 2009; Haibo He & Garcia, 2009). Sin embargo se debe tener presente que las muestras desbalanceadas no son el único factor que puede afectar el rendimiento de la RNAs, otros factores tales como el tiempo de entrenamiento y la complejidad de la red pueden afectar la capacidad de predicción del modelo.

De acuerdo con lo sugerido por Gong (1996) los datos de entrada se estandarizaron para que los valores se encontraran en el rango [0,1] ya que esto acelera el entrenamiento de la red neuronal. Además la estandarización permite obtener mejores inicializaciones en los pesos, lo que reduce la posibilidad de encontrar un mínimo local. La transformación de

los datos se realizó mediante el uso de los valores máximos y mínimos en la escala del conjunto de datos original. Al escalar cada variable, la RNA asigna igual importancia a los datos, por lo tanto se evita el problema de variables con diferente relevancia debido a su rango de valores. La transformación se realizó a partir de la siguiente ecuación.

$$x'_i = (x_i - \text{mínimo})/(\text{máximo} - \text{mínimo}) \quad (20)$$

Con la muestra balanceada y estandarizada se entrenaron redes cambiando el valor de los parámetros. Según las recomendaciones de Moreira & Fiesler (1995) los valores considerados para la tasa de entrenamiento fueron: 0.01, 0.05, 0.1, y para el momento: 0.9 y 0.5. Estos valores permiten la evaluación de diferencias en el comportamiento de la red lo que hace que el entrenamiento sea más robusto. Igualmente se varió la cantidad de neuronas en la capa oculta y se tomó como base lo sugerido por Hecht-Nielsen (1987) quien afirma que el uso de  $2n + 1$  neuronas ocultas (con  $n$  como el número de neuronas de entrada) puede garantizar el perfecto ajuste de cualquier función continua. En total se entrenaron 24 redes con un ciclo de entrenamiento de 5000 iteraciones, usando la función de activación logística tanto para la capa oculta como para la capa de salida y se utilizó el método de aprendizaje en línea por su capacidad de trabajar con grandes conjuntos de datos, como se indicó en el marco conceptual.

La estandarización utilizada para el modelo de las RNAs también se utilizó para estimar el modelo de regresión logística, en el cual se utilizó inicialmente un tamaño de muestra de 700.000 y 20.000 puntos, pero también se estimó un modelo de regresión con la muestra balanceada de 250.000 puntos para realizar la comparación de la predicción con el modelo de RNA.

Para modelar la deforestación a partir de RNAs no es necesario realizar un análisis de correlación ya que el método de entrenamiento (on line) utilizado en la presente investigación tienen la capacidad de trabajar con variables correlacionadas. Sin embargo para los modelos que involucren mayor cantidad de variables sería recomendable realizar



un análisis de correlación o preferiblemente de componentes principales, ya que una red con menor número de entradas tiene menor cantidad de parámetros que estimar, lo que facilita el entrenamiento y mejora la capacidad predictiva de la red neuronal.

## 6.7 Validación

Se utilizó el Criterio de Información de Akaike (CIA) para evaluar los modelos estimados a partir de la regresión logística y el área bajo la curva ROC (AUC) para los dos tipos de modelos. Esto permitió identificar el modelo de regresión logística que presentó mejor predicción con diferentes tamaños de muestra, algunas tendencias en el entrenamiento de la red frente a diferentes conjuntos de parámetros, así como las redes neuronales más adecuadas para predecir la deforestación.

Las curvas ROC (Receiver Operating Characteristic en inglés) son índices de la exactitud diagnóstica de una prueba y tienen la capacidad de clasificar los resultados aplicando diferentes umbrales para predecir la pertenencia a una clase, por lo tanto proporcionan un criterio unificador en el proceso de evaluación de un modelo (Fawcett, 2006). Una descripción detallada de la aplicación de la curva ROC para la predicción de cambios en las coberturas terrestres se encuentra en (Pontius Jr & Batchu, 2003) y para realizar comparaciones entre un modelo de red neuronal y un modelo de regresión logística en (King 2003).

En la presente investigación el gráfico de la curva ROC se obtiene al definir varios valores umbrales para considerar un pixel como deforestado y graficar la tasa de verdaderos positivos o sensibilidad contra la tasa de falsos positivos o **1- especificidad**.

La tasa de verdaderos positivos y falsos positivos de un clasificador está definida como:

Tvp: tasa de verdaderos positivos

Tfp: tasa de falsos positivos

$$Tvp = \frac{vp}{total\ positivos}$$

$$Tfp = \frac{fp}{total\ negativos}$$

**Vp (verdadero positivo):** Cantidad de pixeles correctamente clasificados como deforestados.

**Fp (falso positivo):** Cantidad de pixeles incorrectamente clasificados como deforestados.

Los términos asociados con la curva ROC pueden definirse como:

**Sensibilidad** = Tvp

**1- Especificidad** =Tfp

$$Especificidad = \frac{verdaderos\ negativos}{falsos\ positivos + verdaderos\ negativos}$$

**Vn (verdadero negativo):** Cantidad de pixeles correctamente clasificados como no deforestados.

## 6.8 Ensamble de Redes Neuronales

Luego de evaluar el rendimiento de las redes neuronales se eligieron aquellas que presentaron mayores valores de AUC para realizar un comité o ensamble de redes neuronales, ya que las capacidades de aproximación de un conjunto de redes es mucho mejor que usar sólo uno de las redes entrenadas (Rojas 1996a). Este conjunto podría contener redes que tienen diferente cantidad de neuronas ocultas, o redes con la misma arquitectura, incluso podrían incluir diferentes tipos de red neuronal (Bishop 1995).

La función F generada por el comité de redes está dada por.

$$F = \frac{1}{N} \sum_{i=1}^N fi \quad (21)$$

Donde N corresponde al número de redes neuronales consideradas en el comité y  $fi$  la predicción de la i-esima red para  $i=1, \dots, N$

A partir de las estimaciones generadas del ensamble de redes neuronales se elaboró la superficie de probabilidad de deforestación para la zona en bosque en 1980.

El **AUC** del escenario base, estimado a partir del ensamble de redes reflejaría la calidad de la proyección de la superficie de deforestación del escenario simulado.

## 6.9 Estimación del Coeficiente de Kappa

Dado que el mapa generado por el ensamble de redes neuronales no proporciona información acerca de la cantidad de deforestación, se estableció un umbral para clasificar los valores continuos en valores categóricos (deforestado=1, no deforestado=0). Para esto se evaluó el coeficiente de Kappa para diferentes umbrales con un incremento de 0.05, lo que permitió evaluar la proporción de acuerdo entre el mapa generado por el ensamble y el mapa de deforestación observada para el periodo 1980-2000. El umbral para el cual el coeficiente de Kappa mostró alto grado de concordancia fue el elegido para estimar la cantidad de deforestación en el escenario simulado.

En términos simples, el coeficiente Kappa (K) corresponde a la proporción de concordancias observadas sobre el total de observaciones, una vez excluidas las concordancias atribuibles al azar (Landis & Koch, 1977). El coeficiente toma valores entre -1 y +1, mientras más cercano a +1, mayor es el grado de concordancia, por el contrario, mientras más cercano a -1, menor grado de concordancia.

En términos conceptuales, la fórmula del coeficiente Kappa puede expresarse como se indica en la siguiente ecuación:

$$k = \frac{[(\sum \text{concordancias observadas}) - (\sum \text{Concordancias atribuibles al azar})]}{[(\text{total de obsevaciones}) - (\sum \text{concordancias atribuibles al azar})]}$$

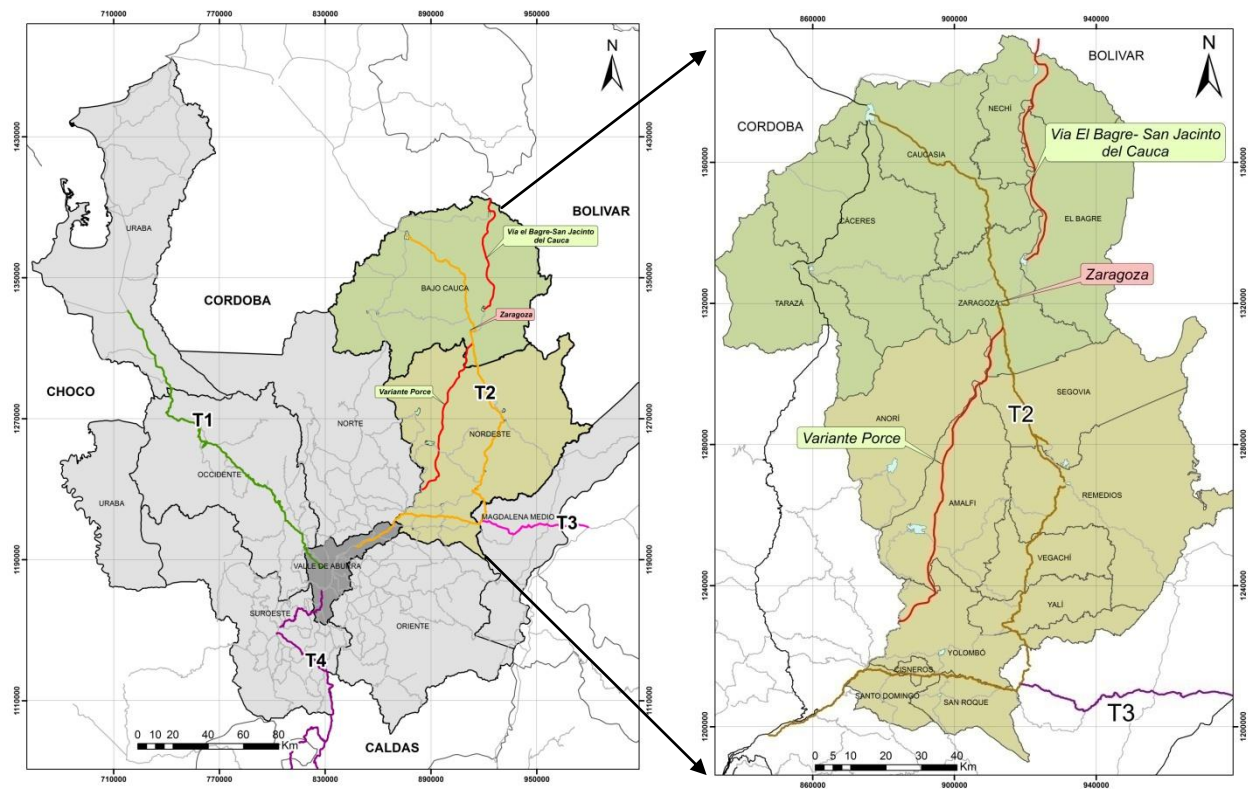
En la tabla 1 se muestra la fuerza de concordancia de acuerdo con los coeficientes de Kappa .

**Tabla 1.** Valoración del coeficiente kappa (Landis & Koch, 1977).

<b>Coeficiente de Kappa</b>	<b>Fuerza de la concordancia</b>
0	Pobre
0,01-0,2	Leve
0,21-0,40	Aceptable
0,41-0,60	Moderada
0,61-0,80	Considerable
0,81-1,00	Casi Perfecta

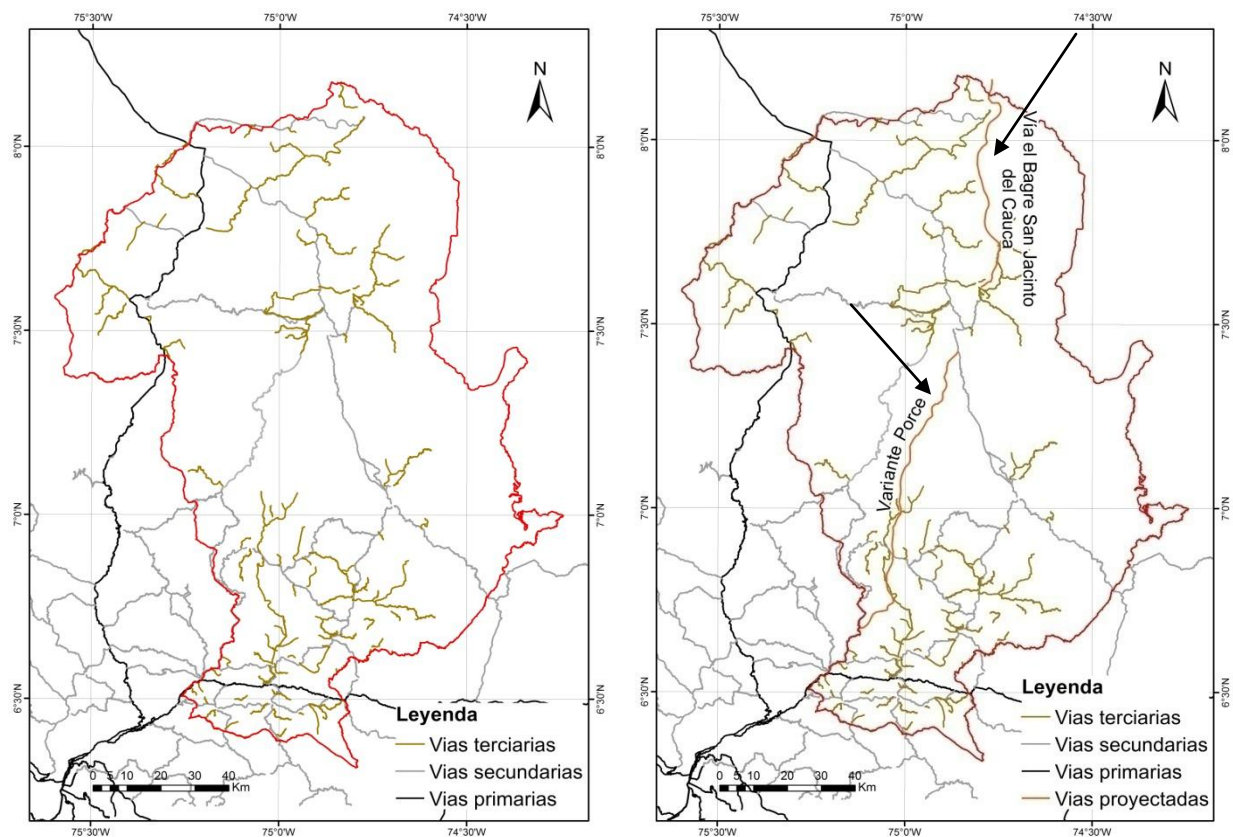
#### **6.10 Simulación del Efecto de la Construcción de Nuevas Vías.**

El escenario base se obtiene al realizar el ensamble de redes con los valores originales de las variables. Para el escenario simulado se incorporaron nuevas vías como la vía El Bagre - San Jacinto del Cauca, de 72 Km de longitud, que se planeó construir para mejorar la conectividad del departamento de Antioquia con el norte del país, y la Variante Porce que va desde el municipio de Yolombó hasta el municipio de Zaragoza (Secretaría de Infraestructura Física 2009). Esta vía de aproximadamente 150 km de longitud se planeó construir en la región de estudio para reducir el tiempo de viaje desde la ciudad de Medellín hacia la costa. Los estudios de esta vía fueron objetados por la autoridad ambiental en la década de los 90 por razón de impactos ambientales sobre ecosistemas, sin embargo en la actualidad se busca retomar los estudios ya que esta variante llegaría a Zaragoza con unos 60 km menos de recorrido que lo que se propone con el tramo 2 (T2) de las Autopistas de la Prosperidad (figura 10).



**Figura 10.** Vías proyectadas para la región de estudio.

Al incluir las nuevas vías (figura 11b) se actualizó la variable distancia a las vías, la variable distancia al borde del bosque y la variable distancia de menor costo a las cabeceras municipales, para lo cual se actualizó el valor de impedancia por tipo de cobertura y tipo de vía. Luego se estimó de nuevo el ensamble de redes manteniendo los valores originales de las otras variables explicadoras, para finalmente obtener la superficie de probabilidad de deforestación del escenario simulado. En la figura 11 se presenta la red vial utilizada para el escenario base y el escenario simulado.



**Figura 11.** Red vial empleada para la construcción de las variables explicadoras. **a)** Escenario base. **b)** Escenario simulado.

Los pixeles afectados por las nuevas vías se identificaron al superponer los archivos raster de la deforestación estimada en el escenario base con la deforestación estimada en el escenario simulado. La deforestación neta adicional corresponde al área representada por aquellos pixeles que se predicen como no deforestados en el escenario base pero que si se deforestaran en el escenario simulado.

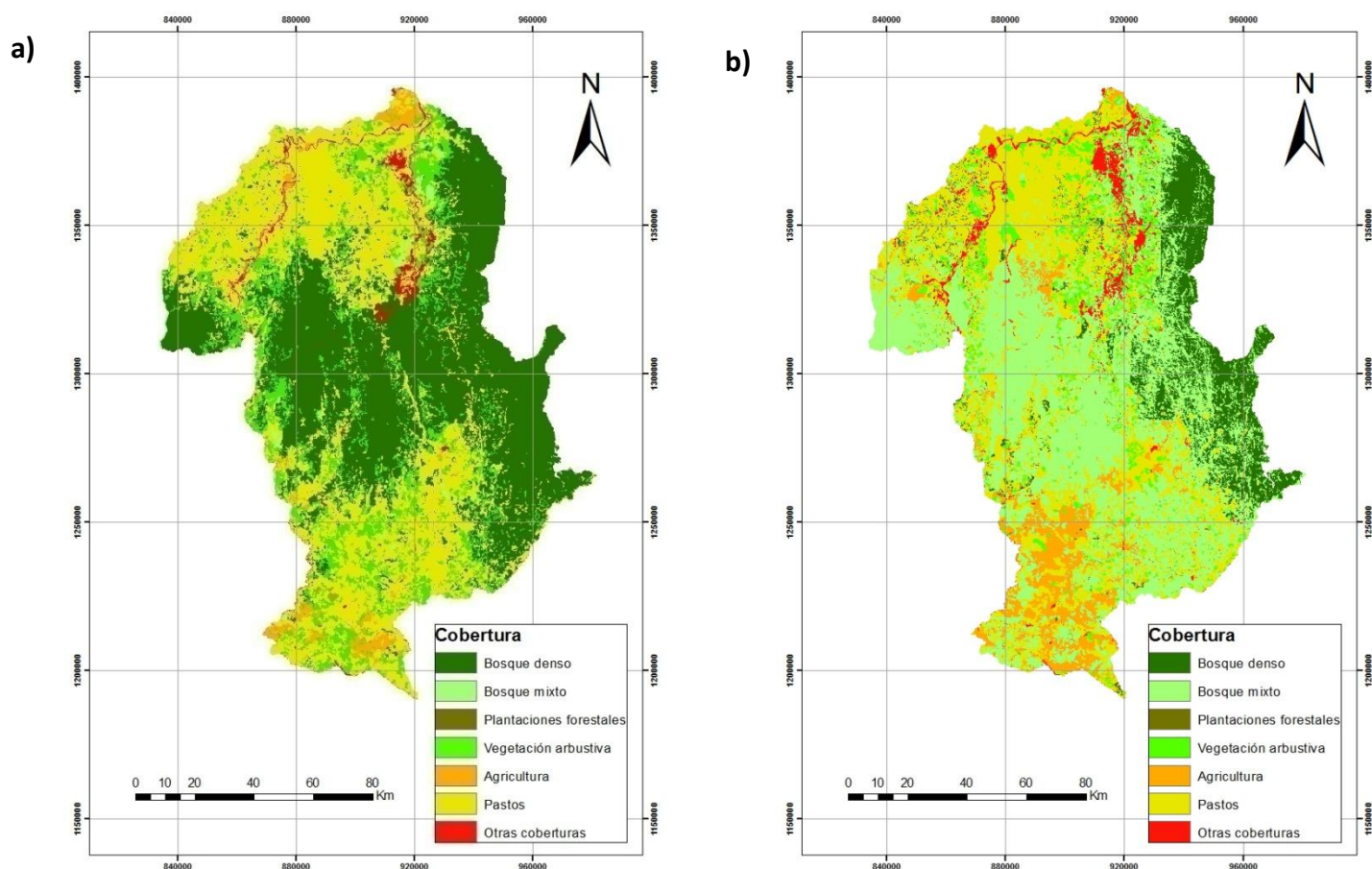
## **7. RESULTADOS**

En la Tabla 2 se presenta la matriz de cambios en coberturas terrestres para el área de estudio y el período 1980-2000. Los cambios corresponden a cálculos obtenidos a partir de mapas de coberturas terrestres proporcionados por Orrego (2009). En cuanto a la magnitud de los cambios, el bosque se convirtió principalmente a pastos y a Bosque Mixto. Durante el periodo de análisis, algunas tierras en cultivos se cubrieron con vegetación arbórea de porte bajo. Otras tierras en pastos o con vegetación arbórea de porte bajo pasaron a bosques, en una transición que caracteriza la recuperación de la cobertura forestal. En el área de estudio se observó un proceso significativo de degradación forestal, al convertirse 58% de la superficie original de bosques densos a bosques intervenidos a finales de la década de los noventa (tabla 2). En la figura 12, se observa las coberturas terrestres con una matriz del paisaje dominada por bosques mixtos y pastos para el año 2000.

**Tabla 2.** Matriz de cambios en coberturas terrestres, 1980-2000. Valores en hectáreas

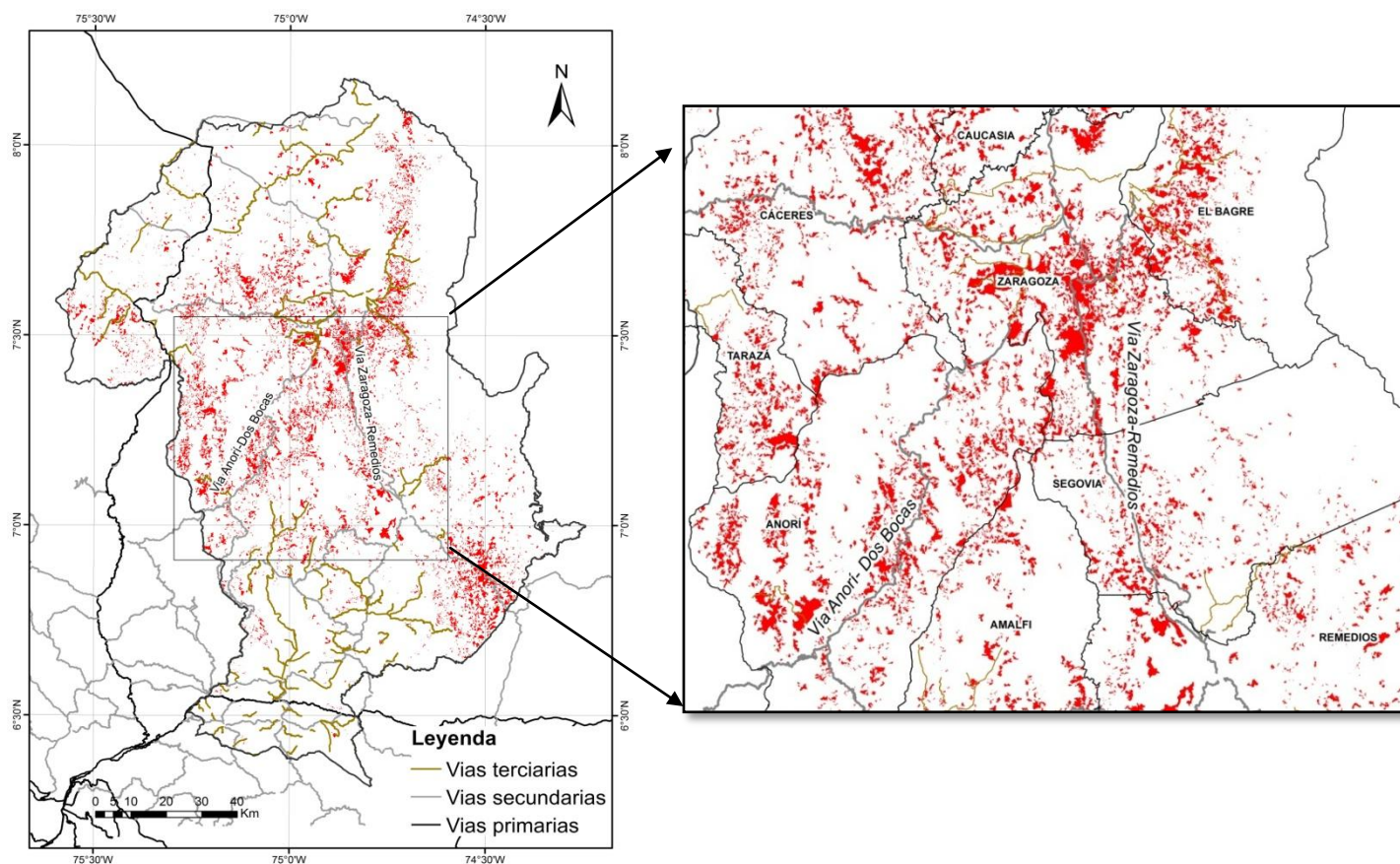
Cobertura terrestre, 1980	Cobertura terrestre, 2000							Total
	Bosque Denso	Bosque Mixto	Plantaciones forestales	Vegetación Arbustiva	Agricultura	Pasto	Otras Coberturas	
Bosque Denso	161,157.92	398,858.78	14.64	27,519.09	10,268.64	79,310.19	5,822.90	682,952.15
Bosque Mixto	131.71	6,171.64	3.25	2,085.37	440.02	4,546.86	599.30	13,978.16
Plantaciones forestales	0.00	375.86	6.25	29.50	593.89	500.60	1.10	1,507.19
Vegetación Arbustiva	5,864.34	100,270.63	74.29	26,923.29	29,657.12	78,955.56	9,042.06	250,787.29
Agricultura	458.86	5,120.68	0.07	1,317.04	11,290.68	11,846.63	1,381.17	31,415.14
Pastos	14,986.02	159,631.38	256.25	53,436.97	63,676.70	287,955.24	24,668.06	604,610.63
Otras coberturas	367.60	5,384.54	5.29	3,527.90	858.20	8,243.41	12,743.98	31,130.93
<b>Total</b>	182,966.44	675,813.52	360.04	114,839.16	116,785.25	471,358.49	54,258.58	1,616,381.49





**Figura 12.** Coberturas Terrestres de las Regiones del Nordeste y Bajo Cauca, Antioquia. **a)** 1980, **b)** 2000.

La deforestación en el área de estudio para el periodo 1980-2000 fue de aproximadamente 130.610ha, ésta cantidad corresponde a la suma de los valores que se presentan en el cuadro rojo de la tabla 2, es decir a la cantidad de cobertura forestal de 1980 (bosque denso, bosque mixto) que para el año 2000 paso a otro tipo de coberturas. La deforestación se presentó con una tasa de 0.0108% anual y ocurrió principalmente en los municipios de Remedios, Zaragoza, y Anorí, con 20.700 ha, 18.943 ha y 18.327 ha deforestadas respectivamente. La deforestación se observa a lo largo de la vía Anorí Dos Bocas, y cerca de las cabeceras de los municipios del Bagre y Zaragoza a lo largo de las vías terciarias y la vía Zaragoza-Remedios (figura 13). La mayoría de los bosques se deforestaron para establecer pastos cerca a las vías.



**Figura 13.** Deforestación observada en las regiones del Nordeste y Bajo Cauca, Antioquia, 1980-2000. Tamaño de pixel: 30 m.

### 7.1 Modelo de regresión logística y validación

Para todos los tipos de muestras, el análisis de correlación arrojó un valor superior a 0.6 en un par de variables: distancia a vías y distancia de menor costo a las cabeceras municipales (ANEXO A). Por lo tanto, se estimó un modelo excluyendo la variable distancia de menor costo y otro incluyendo todas las variables explicadoras (tabla 3, 4 y 5).

Los modelos con menor CIA y mejor rendimiento, es decir un AUC mayor, fueron los modelos estimados con todas las variables explicadoras, nombrados como M1, M3 y M5, presentados en las tablas 3, 4 y 5 respectivamente. Los otros modelos (M2, M4 y M6) presentaron predicciones inferiores pero muy similares a los modelos de mejor

rendimiento, por lo tanto la exclusión o inclusión de la variable distancia de menor costo no generó cambios significativos en la predicción. Aunque en los modelos M1 y M3 los cuales se estimaron con todas las variables y con muestras en las cuales no se tuvo en cuenta el posible efecto de la autocorrelación espacial, estos presentaron mejor capacidad predictiva con respecto al modelo M5 lo que puede ser explicado por la diferencia en el tamaño de la muestra. Para todos los modelos, la magnitud de los parámetros de las variables distancia a vías y distancia al borde del bosque indica la importancia de estos factores sobre la deforestación presente durante el periodo 1980-2000.

En las tablas 3, 4 y 5 se presentan los coeficientes de regresión de los modelos estimados con un tamaño de muestra de 700,000, 250,000 y 20,000 puntos respectivamente.

**Tabla 3.** Coeficientes de regresión estimados para una muestra de 700.000 puntos.

<b>Variables explicadoras</b>	<b>Coeficientes de Regresión</b>	
	<b>M1</b>	<b>M2</b>
Intercepto	0.0965***	-0.1410***
Distancia a ríos	-0.2561***	-0.0738***
Distancia a vías	<b>-3.6448***</b>	<b>-4.5801***</b>
Distancia de menor costo a las cabeceras municipales	-1.3139***	_____
Distancia al borde del bosque	<b>-6.8115***</b>	<b>-7.0798***</b>
Aptitud agrícola	0.7274***	0.7800***
Pendiente	-1.4224***	-1.6442***
<b>CIA</b>	5181.60	5202.43
<b>AUC</b>	0.7760	0.7743

Valores de significancia:  $p < 0.000$  \*\*\*

**Tabla 4.** Coeficientes de regresión estimados para una muestra balanceada de 250.000.

<b>Variables explicadoras</b>	<b>Coeficientes de Regresión</b>	
	<b>M3</b>	<b>M4</b>
Intercepto	1.6814***	1.4709***
Distancia a ríos	-0.3082***	-0.1164***
Distancia a vías	<b>-3.7894***</b>	<b>-4.6243***</b>
Distancia de menor costo a las cabeceras municipales	-1.1469***	_____
Distancia al borde del bosque	<b>-6.1675***</b>	<b>-6.3463***</b>
Aptitud agrícola	0.8043***	0.8658***
Pendiente	-1.3676***	-1.5897***
<b>CIA</b>	2831.59	2841.67
<b>AUC</b>	0.7755	0.7736

Valores de significancia:  $p < 0.000$  \*\*\***Tabla 5.** Coeficientes de regresión estimados para una muestra de 20.000 puntos.

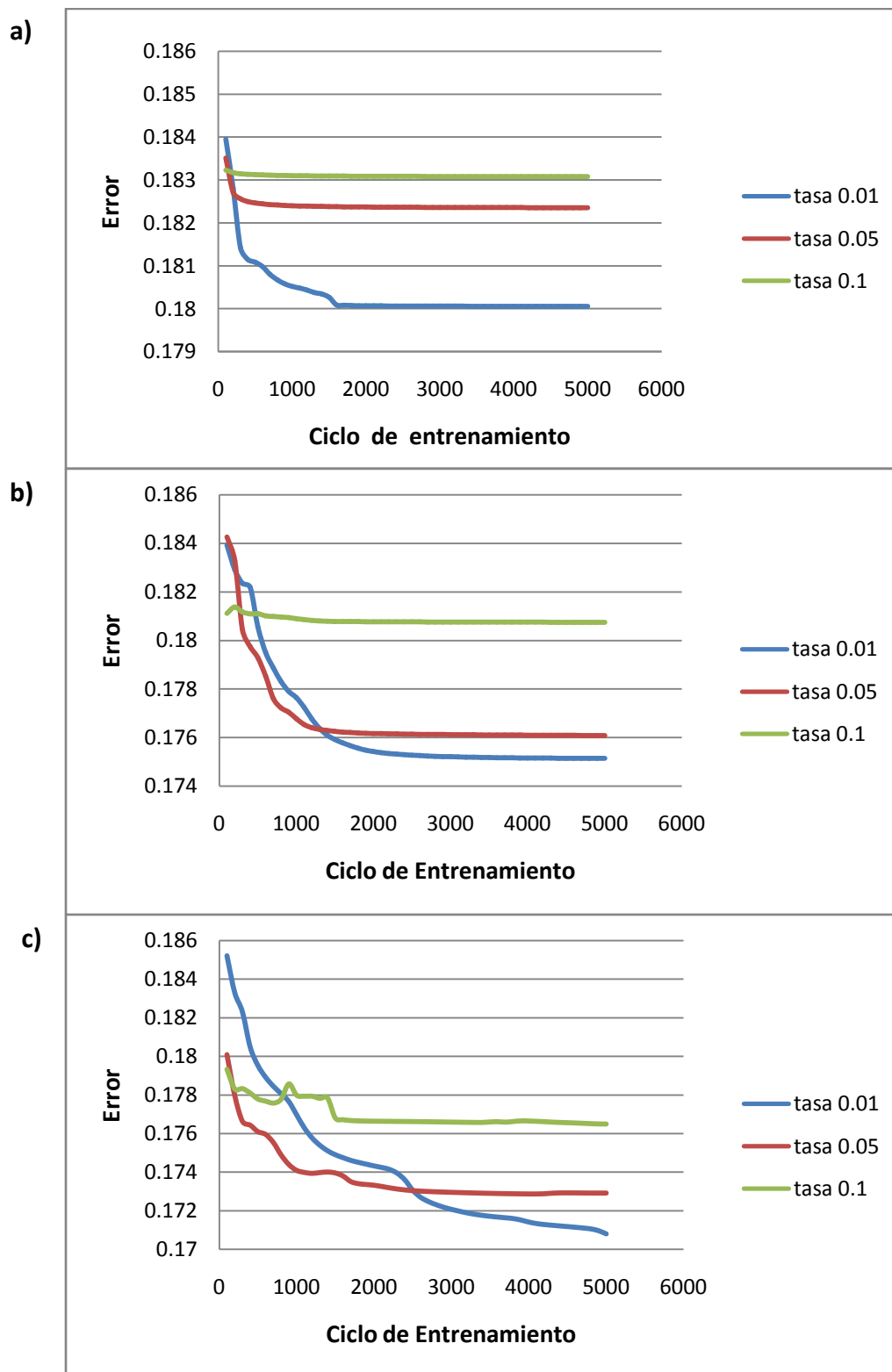
<b>Variables explicadoras</b>	<b>Coeficientes de Regresión</b>	
	<b>M5</b>	<b>M6</b>
Intercepto	-0.8471***	-1.0672***
Distancia a ríos	-0.4654**	-0.2953*
Distancia a vías	<b>-3.2876***</b>	<b>-4.1427***</b>
Distancia de menor costo a las cabeceras municipales	-1.2146***	_____
Distancia al borde del bosque	<b>-2.1246***</b>	<b>-2.3440***</b>
Aptitud agrícola	0.0172*	0.0484*
Pendiente	-0.6097**	-0.8180***
<b>CIA</b>	125.91	126.28
<b>AUC</b>	0.7601	0.7586

Valores de significancia:  $p < 0.01$  \*;  $p < 0.001$  \*\*;  $p < 0.000$  \*\*\*

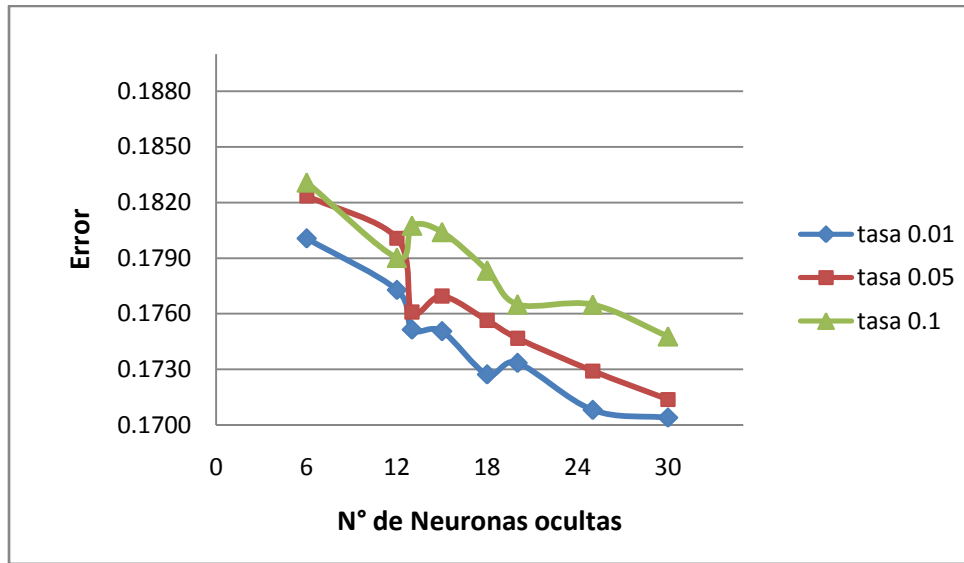
## **7.2 Modelo de red Neuronal y validación**

Durante la fase de entrenamiento de 5,000 iteraciones se obtuvieron menores errores para las redes con mayor cantidad de neuronas en la capa oculta y con tasas de entrenamiento de 0.01 y 0.05, por el contrario para una tasa de 0.1 y menor cantidad de neuronas ocultas la reducción del error fue mínima (figura 14).

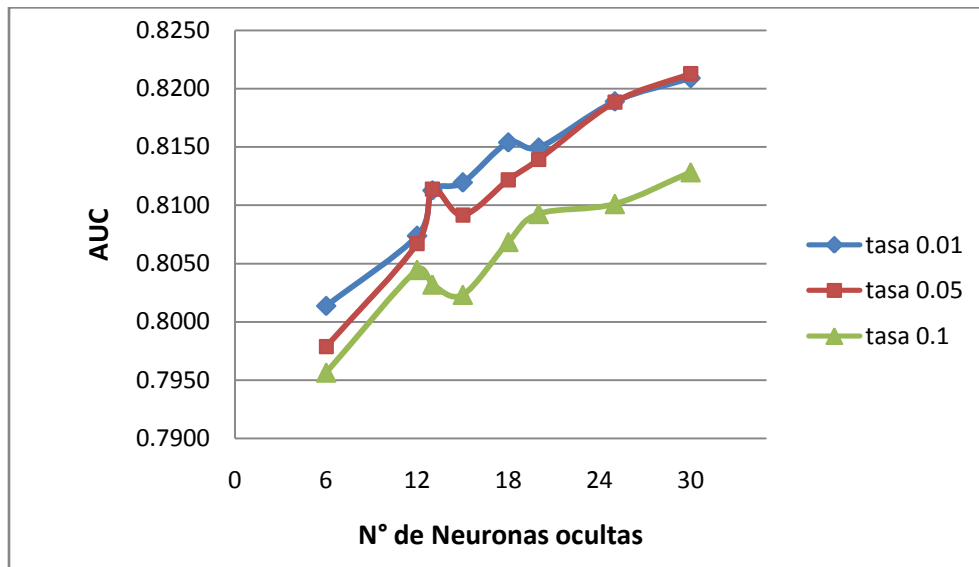
De igual forma las redes que presentaron altos valores de AUC para el conjunto de datos total fueron aquellas que se entrenaron con una menor tasa de entrenamiento y mayor cantidad de neuronas en la capa oculta como se observa en las figuras 15 y 16. En el Anexo B se muestran los parámetros utilizados para el entrenamiento de las 24 redes neuronales, cada una con 6 neuronas en la capa de entrada correspondientes a las 6 variables explicadoras, y una neurona en la capa de salida, así como los valores de AUC los cuales indican la capacidad de generalización de cada red.



**Figura 14.** Reducción del error para un ciclo de entrenamiento de 5,000 iteraciones. **a)** red con 6 neuronas ocultas, **b)** red con 13 neuronas ocultas y **c)** red con 25 neuronas ocultas.



**Figura 15.** Comportamiento del error con respecto a las tasas de entrenamiento y al número de neuronas ocultas para un ciclo de 5000 iteraciones.



**Figura 16.** Área bajo la curva ROC (AUC), con respecto a las tasas de entrenamiento y al número de neuronas ocultas para un ciclo de 5000 iteraciones.

### 7.3 Ensamble de redes neuronales

Con respecto a las 24 redes entrenadas (Anexo B), las 5 redes que presentaron mayor capacidad de generalización se presentan en la tabla 6. Al realizar el ensamble con las 5 redes se obtuvo un AUC de 0.8251, que es un poco superior a los AUC de las redes consideradas en el ensamble.

**Tabla 6.** Parámetros de las redes neuronales artificiales que presentaron mejor capacidad predictiva.

RED	Momento	Tasa de entrenamiento	Número de neuronas ocultas	Error Cuadrático medio	AUC
					Validación
16	0.5	0.01	20	0.173348	0.814942
19	0.5	0.01	25	0.170802	0.818888
20	0.5	0.05	25	0.172912	0.818829
22	0.5	0.01	30	0.170392	0.820893
23	0.5	0.05	30	0.171380	0.821269
					<b>0.8251631</b>

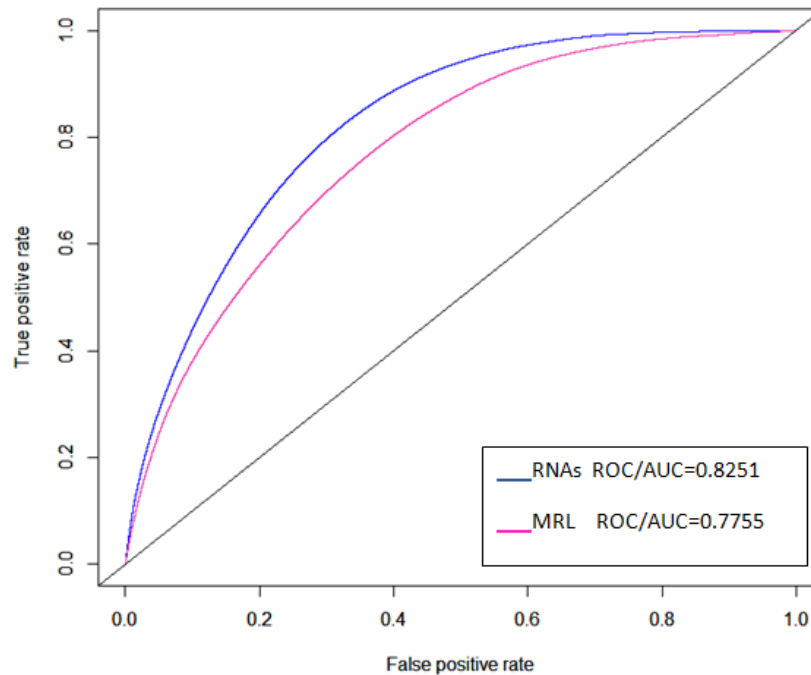
### 7.4 Comparación entre la regresión logística y RNAs (AUC/ROC)

La curva ROC es un gráfico en el que se observan todos los pares Tvp/Tfp resultantes de la variación continua de los puntos de corte (umbrales) en todo el rango de resultados observados.

La figura 17 muestra una línea negra (diagonal) la cual representa una capacidad de predicción del 50% que corresponde a lo esperado por el azar para un variable dicotómica y la diferencia entre la curva ROC obtenida para el modelo 3 estimado a partir de la regresión logística (línea morada) y la curva ROC del ensamble de RNAs (línea azul), así como los valores de AUC. Ambas curvas corresponden a los modelos estimados con la muestra balanceada de 250,000 puntos.

El AUC del ensamble de redes es mayor que el AUC de la regresión logística, por lo tanto hay mayor proporción de aciertos (Tvp) que desaciertos (Tfp), lo que indica que las redes neuronales tienen mejor capacidad predictiva con respecto al modelo de regresión logística.



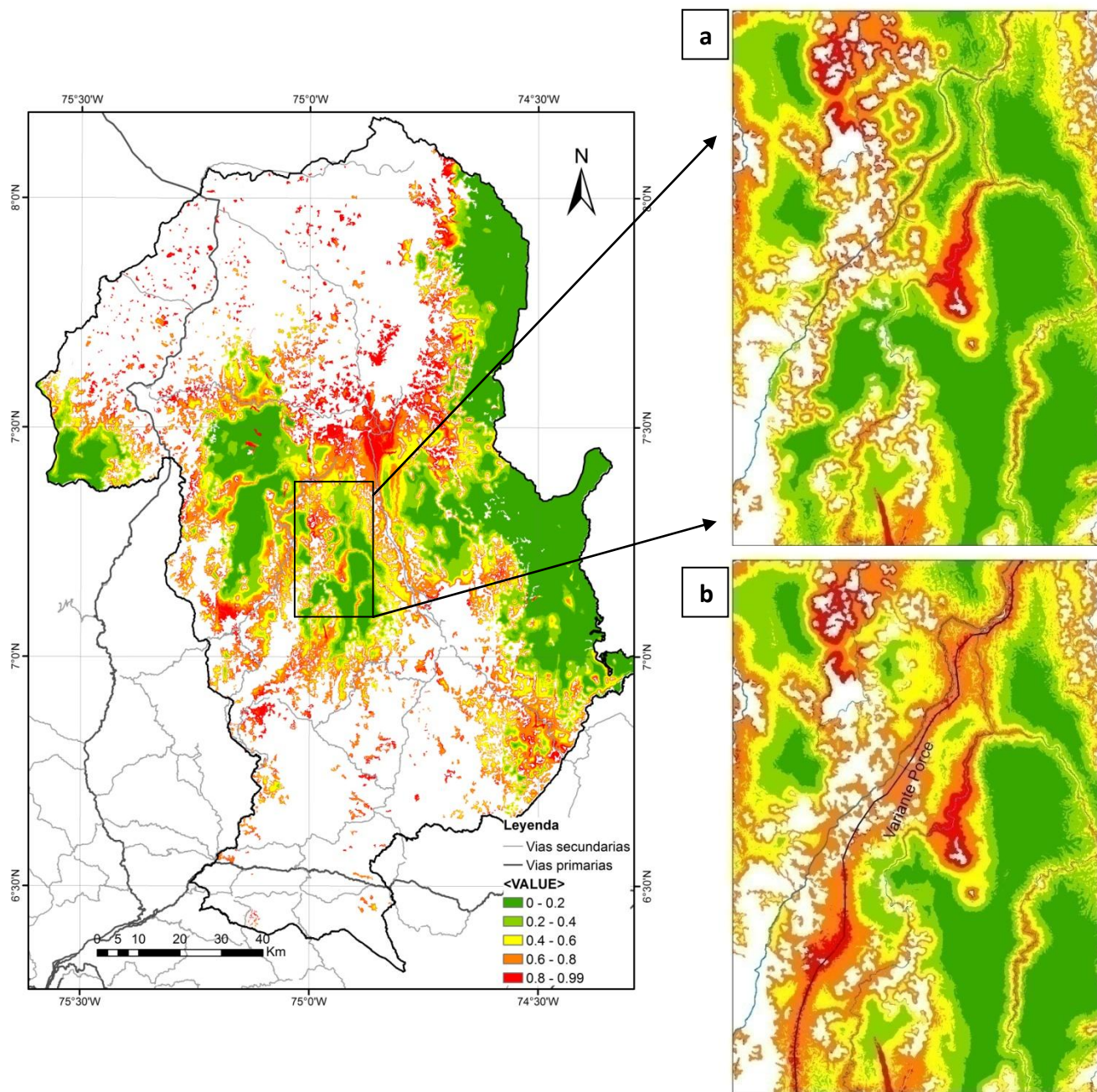


**Figura 17.** Evaluación de la capacidad predictiva entre el ensemble de RNAs y la Regresión Logística (ROC/AUC).

### 7.5 Superficies de Deforestación

A partir de las estimaciones del ensemble de RNAs se generó una superficie de probabilidad de deforestación para el área cubierta por bosques en 1980 (figura 18a). El ensemble predice bajas probabilidades de deforestación (color verde) en la parte oriental de la región de estudio. Estas constituyen áreas remotas, con altas pendientes, baja densidad vial, donde se observó la menor conversión de bosques naturales entre 1980 y 2000. Las altas probabilidades de deforestación (en color rojo) se presentan a lo largo del borde del bosque y en la parte central cerca de la cabecera municipal de Zaragoza donde hay mayor densidad vial.

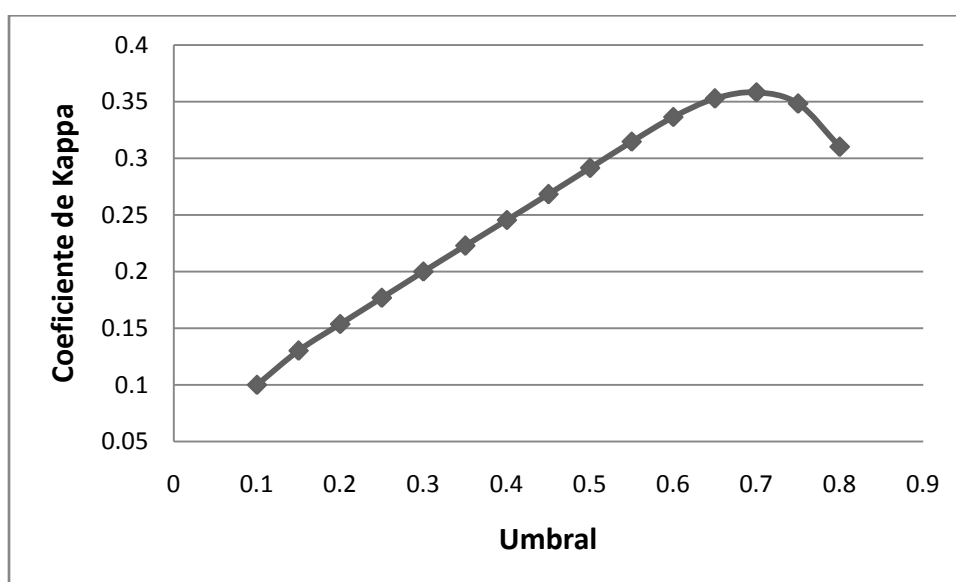
Al igual que el escenario base, se generó una superficie de probabilidad de deforestación para el escenario simulado a partir del ensemble de las RNAs (figura 18b) y se observa que las zonas de mayor potencial para la deforestación son visibles en los bordes de los segmentos boscosos y a lo largo de las nuevas vías.



**Figura 18.** Superficie de probabilidad de deforestación para las regiones del Bajo Cauca y Nordeste, Antioquia. **a)** Escenario base **b)** Escenario simulado. Tamaño de pixel 30m.

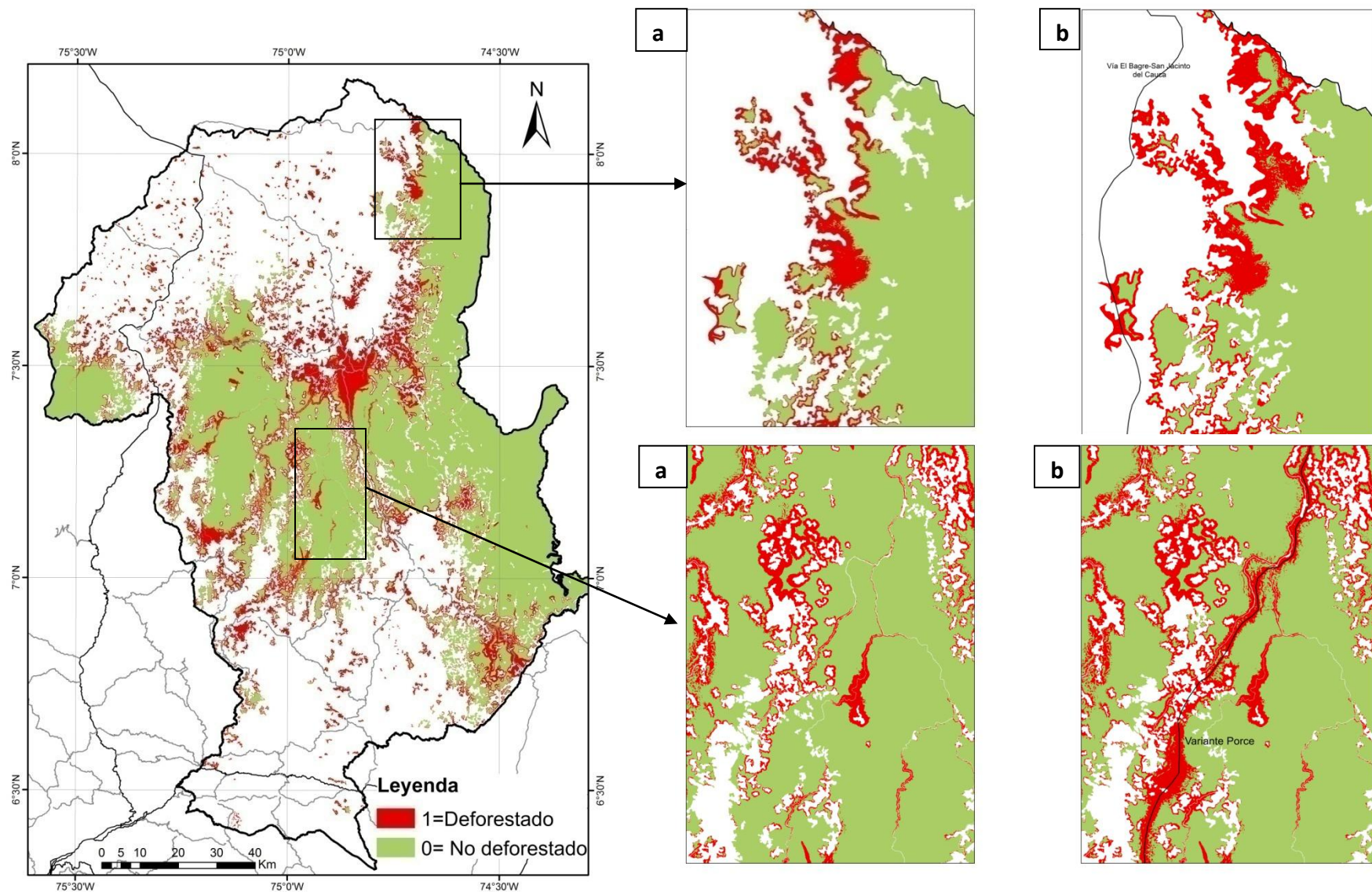
## 7.6 Coeficiente de Kappa

Se eligió un umbral de 0,7 para considerar un pixel como deforestado. Con este umbral el coeficiente de Kappa fue de 0,35 (figura 19) lo que indica que la concordancia de la deforestación estimada por el ensamble de RNAs (figura 20a), con la deforestación observada durante el período 1980-2000 se encuentra dentro de un rango aceptable (tabla 1). Por lo tanto este umbral fue el utilizado para estimar la deforestación en el escenario simulado (figura 19b).



**Figura 19.** Coeficiente de Kappa para diferentes umbrales.



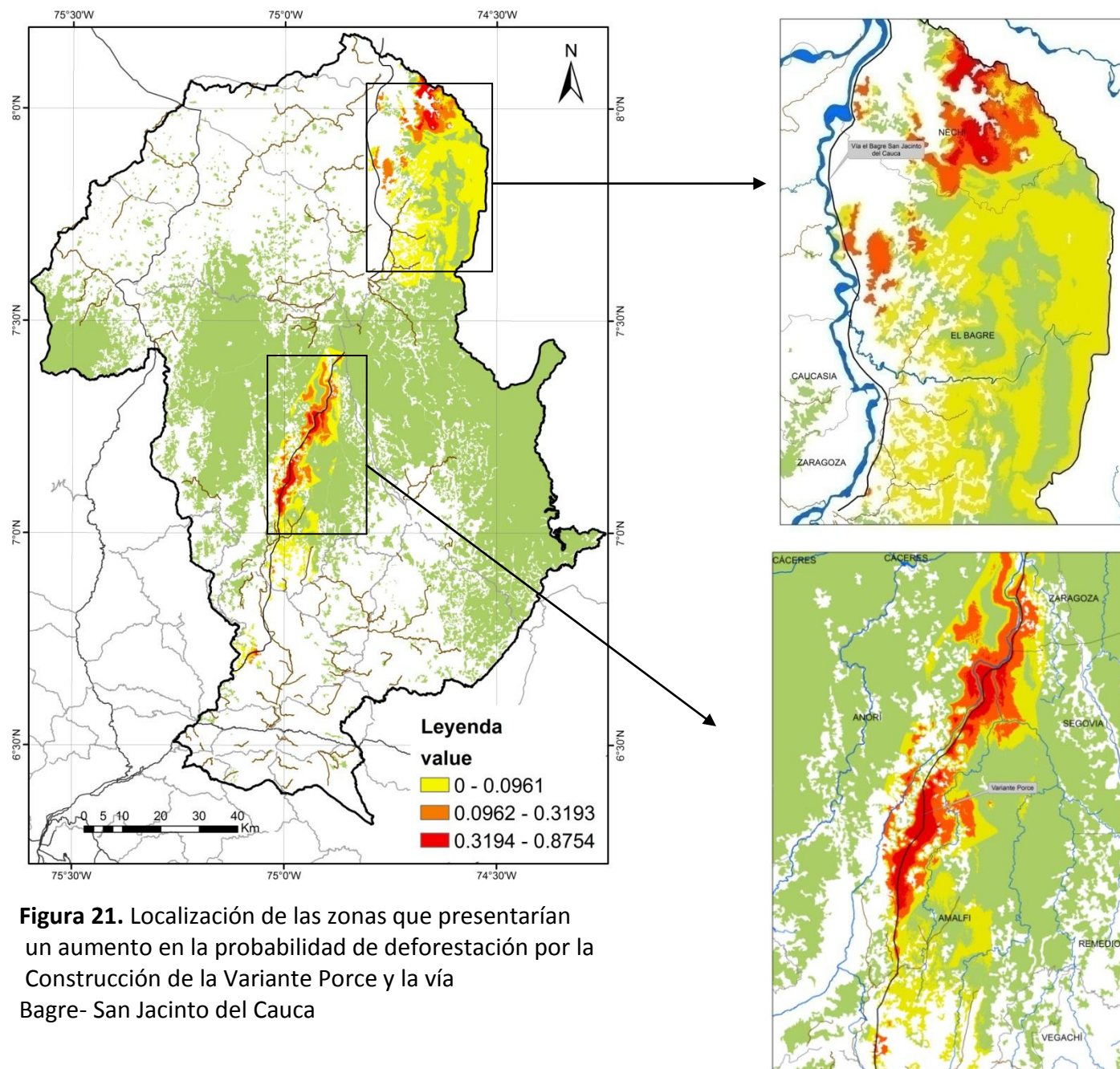


**Figura 20.** Deforestación estimada para las regiones del Bajo Cauca y Nordeste, Antioquia. **a)** Escenario base **b)** Escenario simulado. Probabilidad que un pixel se considere deforestado: 0.7

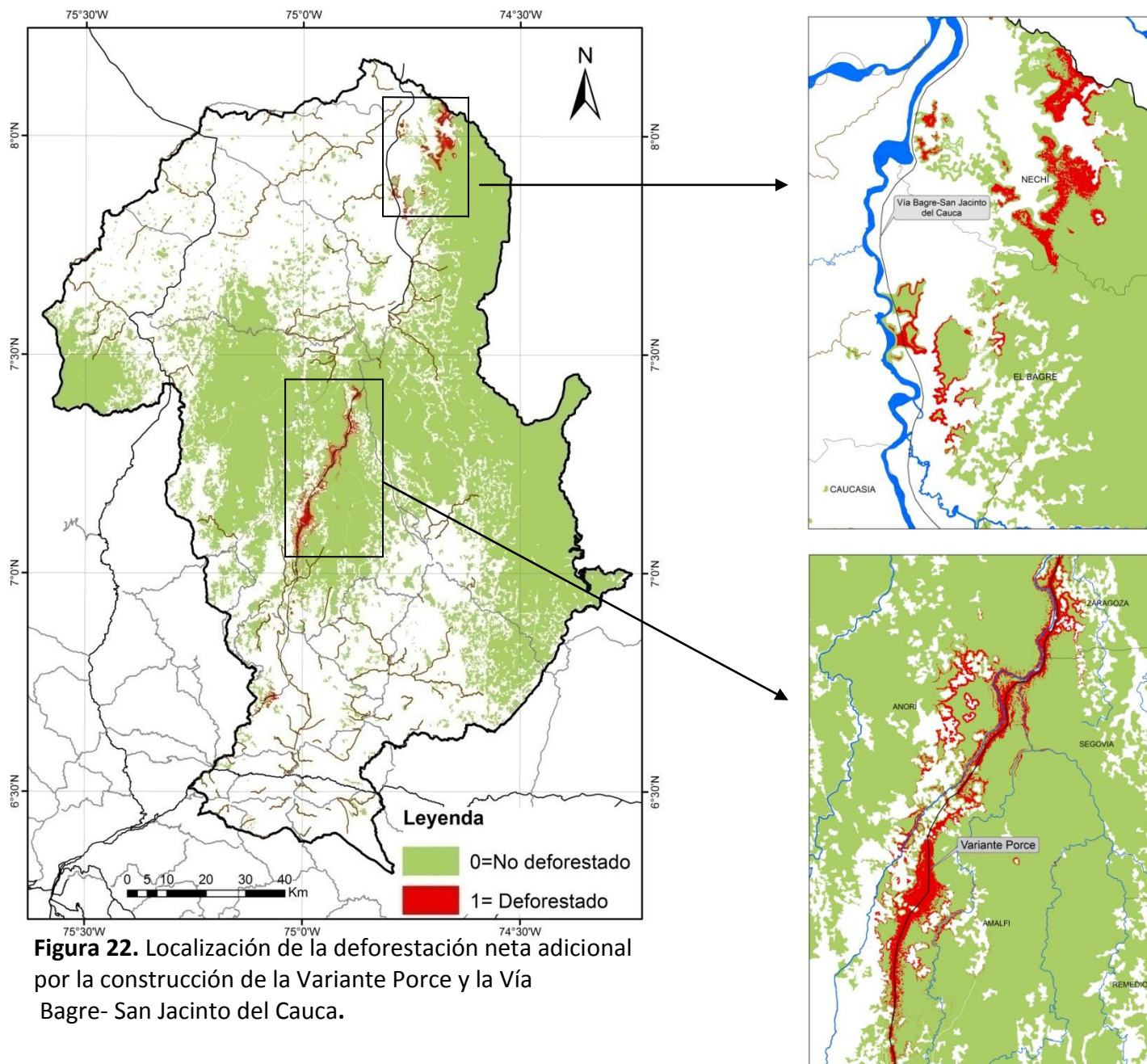
## **7. 7 Simulación del efecto de la construcción de nuevas vías: Variante Porce, Vía Bagre-San Jacinto del Cauca.**

La evaluación del efecto potencial de la construcción de las nuevas vías sobre la deforestación mediante la comparación de un escenario base y otro simulado, sugiere que 1,136, 997 pixeles, aproximadamente 102.329ha (área de pixel de 900 m<sup>2</sup>) incrementarían su probabilidad de deforestación (figura 21), de las cuales 74,050ha presentarían un incremento bajo (color amarillo), 21,471ha un incremento medio (color naranja) y 6,808ha un incremento alto (color rojo). Sin embargo algunos pixeles aproximadamente 7% presentaron reducción en la probabilidad de deforestación para el escenario simulado lo que va en contra vía de lo esperado según la literatura, que indica que la probabilidad de deforestación incrementa con la cercanía a las vías (Chomitz & Gray, 1996; Nelson & Hellerstein, 1997). Esto puede ser atribuible al error esperado por la simulación si se tiene en cuenta que la capacidad predictiva del ensamble de redes fue del 82%, por lo tanto a estos pixeles se les asignó un valor de cero en el incremento de la probabilidad de deforestación, es decir que no se consideró ningún cambio para la estimación final.

Al tener en cuenta el umbral de 0,7 la comparación de escenarios indica que la construcción de las nuevas vías probablemente conduciría a una deforestación neta de 119,800 pixeles, aproximadamente 10,782ha (figura 22), de las cuales 4,163ha estarían asociadas a la vía El Bagre-San Jacinto del Cauca y 6,619ha a la variante Porce.







## **8. DISCUSIÓN**

Las regiones del Bajo Cauca y Nordeste del departamento, tenían al inicio del período de estudio una cantidad significativa de bosque denso, aproximadamente 682,952.15ha de las cuales 58% pasaron a bosques mixtos para el año 2000. Por lo tanto la cobertura forestal en casi toda la extensión del área de estudio, corresponde principalmente a cobertura de bosque mixto. Esto se debe a procesos de degradación forestal, generados probablemente por aprovechamientos forestales selectivos en los cuales se extraen las especies forestales con mayor valor comercial.

Aunque los bosques en estas regiones se hayan convertido a cultivos comerciales ilícitos (coca) es difícil tener una estimación fiable de la trayectoria de la cubierta vegetal. Sin embargo, algunas estimaciones publicadas indican que alrededor de 2,500 hectáreas fueron cultivadas con coca en el año 2000 (López & Blanco, 2008), lo que representa sólo el 0,04% de la superficie del departamento de Antioquia y 0.15% del área de estudio.

### **8.1 Modelo de Regresión Logística**

Los signos de los parámetros de las variables coinciden con los signos esperados de manera anticipada. Una mayor distancia a vías, ríos y a las cabeceras municipales, así como una mayor pendiente conllevan a una menor deforestación. Por el contrario la aptitud agrícola presenta un efecto positivo sobre la deforestación. Es decir, aquellas áreas más aptas para la agricultura tienen mayores tasas de conversión de bosques, lo que es consistente con los resultados de un estudio de deforestación en Santa Cruz, Bolivia (Mertens et al., 2004).

El valor de los parámetros varió con el tamaño de la muestra utilizado en la estimación de los modelos. No obstante en todos los tamaños de muestra analizados no se observaron cambios en la importancia relativa de las variables explicadoras, definida según la magnitud de los coeficientes de regresión. De acuerdo con los resultados, las variables distancia a vías y distancia al borde del bosque fueron las más importantes en la



deforestación presente durante el período 1980-2000. Estudios realizados a nivel regional en Antioquia también concluyeron que la distancia a las vías fue un determinante de la deforestación (Orrego 2009). Estas variables han sido importantes impulsores del cambio de la cobertura boscosa en el sur de Camerún (Mertens & Lambin, 1997), en las tierras bajas de Sumatra, Indonesia (Linkie et al., 2004) y factores importantes del cambio de los bosques en el sureste de México (Mas et al., 2004).

El análisis de la potencial presencia de multicolinealidad entre variables explicadoras conllevó a analizar los coeficientes de correlación entre pares de variables. Aunque la variable distancia de menor costo presentó alta correlación con la variable distancia a las vías, los modelos estimados incluyendo esta variable arrojaron mejores predicciones con un CIA menor y AUC mayor. Considerar esta variable en los modelos de cambios en el uso del suelo es muy importante porque no solo representa un proxy de la accesibilidad a mercados sino que también podría ser un indicador de la presión demográfica (Thies et al., 2012).

## **8.2 Modelo de RNAs**

Las simulaciones dieron lugar a una reducción del error en las primeras 5,000 iteraciones con valores de convergencia de 0.18 a 0.17, después de 5000 iteraciones el entrenamiento fue más inestable y no se presentaron mejoras considerables en el error, por lo tanto se suspendió el entrenamiento en este ciclo sin pérdidas sustanciales en precisión. Este mismo nivel de convergencia fue alcanzado por Müller & Mburu (2009) quienes modelaron la deforestación con redes neuronales utilizando el algoritmo de retropropagación.

Durante el entrenamiento de la red neuronal se identificaron varias tendencias, por ejemplo las redes que tenían  $2n + 1$  ( $n$ =cantidad de neuronas de entrada) neuronas ocultas o cantidades superiores a esta cifra presentaron altos valores de AUC con respecto a las redes que tenían una cantidad de neuronas por debajo de  $2n+1$ , lo que es consistente con algunos estudios en los cuales recomiendan el uso de  $2n+1$  neuronas en la capa oculta

(Hecht-Nielsen, 1987) para obtener una buena generalización. Yilmaz (2009) obtuvo bajos errores con una estructura 8-17-1, al evaluar la susceptibilidad a los deslizamientos. Gómez-Sanchis et al. (2006) analizaron diferentes modelos de RNAs para predecir la concentración de ozono en diferentes períodos de tiempo, teniendo en cuenta el modo de entrenamiento (por lotes o en línea) y la arquitectura de la red (una o dos capas ocultas, con 5 a 16 neuronas ocultas en cada una). Los mejores modelos neuronales se formaron por 8 neuronas en la capa de entrada y una sola capa oculta con 14 a 16 neuronas, y el método en línea (on line) proporcionó el mejor modelo de red neuronal.

Varios investigadores como Lawrence et al. (1996) han estudiado el comportamiento de diferentes estructuras de red neuronal variando el número de neuronas en la capa oculta, sin embargo no hay un consenso general acerca del número óptimo de neuronas a utilizar, lo que representa una desventaja de este método, por lo tanto para cada caso es importante realizar un análisis exploratorio con el fin de determinar la estructura óptima de la red neuronal. Además de una capa oculta la red neuronal puede trabajar con dos o más capas ocultas y menor cantidad de neuronas por capa, lo que permitiría aproximaciones más eficientes en el sentido de lograr el mismo nivel de precisión con menos pesos y sesgos en total, pero de acuerdo con Rumelhart et al. (1986) rara vez representa una ventaja trabajar con dos o más capas ocultas y actualmente es muy poco lo que se sabe acerca del comportamiento de redes con este tipo de estructura (Bishop, 1995).

A medida que la tasa de entrenamiento disminuyó, aumentó el valor del AUC, pero se debe tener presente que aunque redes con bajas tasas de entrenamiento presentan menor error, el entrenamiento es muy lento comparado con las demás tasas. Por otra parte las redes con mejores predicciones frente a la entrada de nuevos datos no siempre son las que presentan el error más bajo, debido a que la red puede estar sobre-entrenada, por ejemplo durante el proceso de entrenamiento se obtuvieron redes con errores de

0.06 que presentaron baja capacidad de generalización con valores de AUC entre 0.70 y 0.72.

Una tasa de entrenamiento de 0.1 y un momentum de 0.9 se consideran valores estándar para el entrenamiento de una red neuronal, sin embargo con estos valores se presentaron oscilaciones en el error, obteniendo para algunas redes errores altos comparados con los errores presentados al inicio de la fase de entrenamiento. Una posible solución para evitar las oscilaciones es que cada vez que el error se incremente en al menos 1% en comparación con la iteración anterior la tasa global de aprendizaje se puede reducir en un 50% (Moreira & Fiesler, 1995). Por lo tanto el valor de la tasa de aprendizaje debe ser lo suficientemente grande como para permitir un proceso de aprendizaje rápido pero lo suficientemente pequeño como para garantizar su efectividad, para la presente investigación la tasa que presentó buen rendimiento en cuanto a tiempos de entrenamiento y capacidad de generalización fue de 0.05.

El algoritmo de retropropagación (backpropagation) ha sido utilizado en múltiples aplicaciones. Otra alternativa para el entrenamiento sería el uso de una red adaptativa con el algoritmo de retropropagación resiliente, en el cual la tasa cambia durante el proceso de entrenamiento, lo que evita el problema de definir una tasa de entrenamiento global que sea apropiada para todo el proceso de aprendizaje (Günther & Fritsch, 2010).

Los modelos generados en la fase de entrenamiento presentaron precisiones relativamente similares, con valores de AUC entre 0.80 y 0.82. Müller & Mburu (2009) también encontraron resultados similares con diferentes arquitecturas de red neuronal, Yilmaz (2009) obtuvo 0,85 de AUC lo que mostró que el mapa de susceptibilidad a los deslizamientos obtenido a partir de redes neuronales tiene una mejor precisión comparado con otro tipo de modelos y Okwuashi et al. (2012) quienes simulaban cambios en el uso del suelo en Lagos, Nigeria para tres períodos de tiempo obtuvieron valores de AUC entre 0.7574 y 0.7850.

Al realizar el ensamble de redes se obtuvo un AUC de 0.8251 y aunque la diferencia no fue significativa con respecto a las redes consideradas en el comité, la predicción fue superior. Es una práctica común en la aplicación de redes neuronales entrenar diferentes redes para luego seleccionar la mejor, sobre la base de los resultados generados en la validación. Hay dos desventajas con este enfoque, en primer lugar, todo el esfuerzo involucrado en la formación de las redes restantes se desperdicia y en segundo lugar, la generalización para el conjunto de validación tiene un componente aleatorio debido al ruido en los datos, por lo tanto estos inconvenientes se pueden superar mediante la combinación de redes para formar un comité (Perrone & Cooper, 1993), lo que da lugar a mejoras en las predicciones con poco esfuerzo computacional adicional.

Al establecer un umbral de 0.7 se obtuvo un coeficiente de Kappa de 0.35. Wilson et al. (2005a) quien evaluó el coeficiente de Kappa para diferentes umbrales encontró mayor proporción de acuerdo utilizando un umbral de 0.5. Mas et al. (2004) obtuvieron un coeficiente de Kappa de 0.34 a partir de un modelo de deforestación con red neuronal y Pontius Jr. et al. (2001) obtuvieron un coeficiente entre 0,34 y 0,53 de mapas predictivos derivados de un modelo de deforestación en Costa Rica.

### **8.3 Comparación de modelos**

Aunque el objetivo principal de la regresión logística era identificar la importancia de las variables explicadoras, se evaluó la capacidad predictiva de los modelos hallando el AUC, lo que permitió realizar una comparación general con los modelos estimados a partir de RNAs, las cuales mostraron una capacidad predictiva superior de 0.82 con respecto al modelo regresión logística que presentó un AUC de 0.77. Esto puede ser resultado de la presencia de relaciones no lineales, las cuales no pueden ser detectadas por el modelo de regresión logística. Sanabria (2009) en su estudio realizado en Bolivia también encontró una diferencia significativa en la predicción de la deforestación con un AUC de 0.92 para las RNAs frente a 0,84 para la regresión logística.

Las redes neuronales permitieron mejorar la predicción al pasar de 0.79 a 0.82, por el contrario, la regresión logística aunque mostró menor capacidad predictiva fue útil para entender la importancia relativa de las variables explicadoras sobre la deforestación, lo cual es difícil de analizar a partir de las RNAs. Una desventaja al utilizar la regresión logística en el análisis espacial de datos, es que éste método supone que los datos son estadísticamente independientes y los datos espaciales tienen la tendencia a ser dependientes debido a la proximidad geográfica, por lo que subestima la presencia de autocorrelación espacial y aunque en la selección de una muestra se tuvo en cuenta una distancia mínima entre puntos de 150m, este esquema de muestreo no es suficiente para eliminar el efecto de autocorrelación. Por lo tanto para identificar los principales determinantes del cambio en el uso del suelo se podrían realizar otros métodos estadísticos como modelos de rezago espacial que tienen en cuenta la autocorrelación a partir de un análisis de vecindad (Overmars et al., 2003).

Aunque el aprendizaje de las redes neuronales es dependiente del tamaño de la muestra, en la presente investigación se observó que incluso para muestras pequeñas de 10,000 y 20,000 datos las redes presentan mayor capacidad predictiva con respecto a la regresión logística, siempre y cuando estas muestras se encuentren balanceadas, por el contrario para muestras desbalanceadas la regresión logística superó en capacidad predictiva a la red neuronal, lo cual indica que el aprendizaje de la red no solo estuvo determinado por el tamaño de la muestra, sino también por el tipo de datos considerados en esta.

#### **8.4 Efecto de la construcción de nuevas vías**

La evaluación del potencial efecto de la construcción de las 2 vías, sugiere un incremento en la probabilidad de deforestación de 102,329 ha con una deforestación neta de 10,872 ha. Wilson et al. (2005a) en su estudio realizado en los bosques templados del sur de Chile encontró un incremento en la probabilidad de deforestación de aproximadamente 27,000ha luego de la construcción de una nueva vía.

La simulación indica solo la cantidad de deforestación asociada a las vías proyectadas. En consecuencia, debe ser una estimación conservadora, ya que muchos senderos y caminos temporales se podrían construir a partir de las nuevas vías facilitando el acceso a los bosques, por lo tanto la deforestación en la región de estudio podría evolucionar hasta convertirse en una franja de más de 10,780 ha deforestadas, lo cual afectaría los servicios ambientales que proveen los ecosistemas forestales.

En Antioquia los bosques en las márgenes de las carreteras se han degradado seriamente por la colonización y establecimiento de potreros como resultado de actividades ganaderas. Por ejemplo la presencia de la carretera al mar entre Medellín y Turbo (Unidad Administrativa Especial del Sistema de Parques Nacionales Naturales de Colombia, 2006) y la apertura del ramal de la carretera panamericana en el tramo Barranquillita y Lomas Aisladas ha incentivado la explotación maderera y la colonización (Ministerio de Ambiente, Vivienda y Desarrollo Territorial et al., 2005)

Freitas et al. (2010) indican que a pesar de que la variable distancia a las vías es un fuerte predictor de la dinámica del bosque en los procesos de deforestación, se detecta su efecto solo cuando la expansión agrícola se ha estabilizado en un determinado período de tiempo. Sin embargo el efecto de la construcción de la Variante Porce que atraviesa zonas de baja aptitud agrícola, podría ser resultado de la minería y de la extracción excesiva de madera y no del establecimiento de cultivos agrícolas, lo que es preocupante si se tiene en cuenta que los estudios ecológicos sugieren que los bosques tropicales con base en suelos pobres tienen niveles más altos de biodiversidad que aquellos ubicados en zonas de menor aptitud agrícola (Huston 1994).

La vía el Bagre San Jacinto del Cauca aunque se encuentra lejos de zonas boscosas, la simulación muestra un incremento en la probabilidad de deforestación en la parte oriental del área de estudio, lo que indica que las vías a pesar de no atravesar zonas boscosas juegan un papel importante en determinar la configuración espacial del paisaje.

Aunque los resultados de la simulación no pueden ser validados con datos reales ya que estas vías aún no se han construido, el desarrollo de este tipo de análisis proporciona una orientación empírica a los planificadores sobre la magnitud y ubicación del daño, por lo tanto una vez finalizados los proyectos viales se necesita una vigilancia estricta por parte de las autoridades ambientales, y el diseño de políticas orientadas a la creación de incentivos y sistemas de compensación para la protección de los bosques, que permita a la población ser partícipe de las actividades de conservación.

## **9. CONCLUSIONES**

Modelar la conversión del bosque puede ser un instrumento importante para la comprensión de la dinámica de la cobertura forestal y ofrece la posibilidad de probar la sensibilidad de los patrones de deforestación a los cambios en las variables seleccionadas. Cualquier modelo queda corto en incorporar todos los aspectos de la realidad, pero éstos proporcionan una valiosa información sobre el comportamiento del sistema bajo un rango de condiciones determinadas.

### **9.1 Resultados**

En la presente investigación se desarrolló un modelo de regresión logística y un modelo de red neuronal a partir de variables biofísicas y de accesibilidad. Los resultados de la regresión indican que las variables, distancia a las vías y distancia al borde del bosque fueron las más importantes en la deforestación presente en 1980-2000. Al comparar los resultados obtenidos con modelos de deforestación previamente estimados para el departamento de Antioquia (Orrego, 2009) y la región de Urabá (Ramírez & Orrego, 2011) se concluye que las condiciones de accesibilidad constituyen significativos determinantes de la conversión de los bosques.

Con diferentes modelos de RNAs en función de su arquitectura y con base en su capacidad de generalización se realizó un comité de redes, lo que permitió obtener estimaciones con precisión y robustez. Las RNAs mostraron una capacidad predictiva superior con valores de 0.79 - 0.82 frente a 0.77 para la regresión logística y a partir de la simulación y la comparación de escenarios se obtuvo información muy importante acerca de la potencial deforestación inducida por la construcción de las nuevas vías en la zona de estudio.

Aunque no existen modelos ideales debido a que el proceso de deforestación depende de factores muy diversos que varían con el tiempo, la investigación presenta una herramienta metodológica diferente a los modelos estadísticos, que puede ser generalizable a otros lugares y que ofrece la posibilidad de producir diferentes escenarios que permitan evaluar las estrategias de gestión del uso del suelo.

Una de las principales limitaciones de ésta investigación, consiste en el uso de información espacialmente explícita para un solo período de tiempo y aunque la construcción de modelos de cambios en el uso del suelo, se basa en el supuesto de que los patrones de deforestación permanecen constantes en el tiempo, es importante contar con información de coberturas para mínimo dos periodos con el fin de tener elementos adicionales que ayuden a evaluar el desempeño del modelo.

## **9.2 Metodología**

Si bien las técnicas estadísticas como la regresión logística y métodos computacionales como las redes neuronales artificiales se han aplicado de forma independiente, la combinación de conocimientos entre estos dos tipos de métodos permitiría a los investigadores diseñar modelos más eficaces que combinen las ventajas de ambos, lo cual proporcionaría un mejor entendimiento de los fenómenos ambientales a partir de la construcción de modelos espaciales.



El desarrollo metodológico de esta investigación proporciona información útil para la comparación de los modelos propuestos con otros que se puedan desarrollar a futuro, lo cual es muy importante si se tiene en cuenta que en los estudios en los que se han aplicado redes neuronales solo presentan una teoría básica a cerca del procedimiento de modelación (Maier & Dandy 2000), es decir que dicha información no se da de forma explícita ya el objetivo principal es mostrar los resultados de la predicción, por lo tanto esta ausencia de información sobre el proceso de modelización, no permite realizar comparaciones significativas entre el rendimiento de diferentes modelos.

Al utilizar RNAs, se encuentra una serie de alternativas en la elección de los parámetros de entrenamiento y en cada etapa del proceso de desarrollo del modelo. En consecuencia, se deben elaborar directrices que ayuden a identificar las circunstancias en las cuales determinada estructura deba ser adoptada. Por lo tanto superar los problemas de la estructura y tamaño de la red sería un gran paso hacia un mayor entendimiento de las redes neuronales artificiales. Además de mejorar el diseño y el análisis a través de métodos formales que no se basen en ensayos de prueba y error y que permitan un uso más productivo y eficaz de las redes en la ecología y ciencias ambientales.

Las RNAs podrían ser una herramienta importante para la toma de decisiones sobre las alternativas de trazado de las estructuras lineales de grandes dimensiones, ya que permite la integración de toda la información ambiental disponible de la zona donde se pretende ubicar la obra civil. Esto ayudaría a técnicos en medio ambiente visualizar cómo se trasformaría el paisaje, conforme se define la ubicación de la infraestructura lineal, lo que representa información valiosa para la formulación y puesta en marcha de proyectos forestales de compensación, bajo el mecanismo de reducción de emisiones por deforestación y degradación, REDD.

Las altas tasas de extracción de productos forestales a menudo se sustentan en la construcción y mejoramiento de vías que facilitan el acceso y fomentan el comercio ilegal de madera. En el pasado ha sido difícil investigar desde una perspectiva geo-espacial los efectos de estos proyectos a partir de métodos computacionales, por la dificultad en el análisis espacial de datos. Sin embargo, los nuevos desarrollos en el software de información geográfica y la rápida disminución de los costos en la capacidad computacional, han facilitado la aplicación de estos métodos y el desarrollo de modelos cuantitativos para mejorar la gestión y conservación de los bosques.

## **10. FUTURAS INVESTIGACIONES**

Hay muchas áreas de investigación que pueden explorarse posteriormente en base a los resultados de esta tesis, para fortalecer la aplicación de las redes neuronales en el campo de la modelación ambiental. Algunas ideas específicas para futuras investigaciones incluyen.

- Explorar alternativas diferentes al algoritmo de retropropagación para el entrenamiento de RNAs.
- Evaluar otros parámetros que influyen en el rendimiento RNAs, como la función de activación, número de capas ocultas, número de iteraciones, tasa de aprendizaje, y otros no abarcados por esta tesis.
- Vincular las RNAs con un modelo Autómata celular con el fin de investigar cómo la disposición espacial de la deforestación varía a través del tiempo y cómo las transiciones podrían estar limitadas por un conjunto de factores ambientales, geográficos y antropogénicos.
- Vincular las RNAs junto con cadenas de markov para el desarrollo de un modelo espacio temporal. Las RNAs permiten la integración de los factores que impulsan el

cambio forestal, mientras que las cadenas de Markov controlan la dinámica temporal de los cambios en la cobertura forestal.

- Complementar los resultados con datos de la distribución espacial de especies. Esto permitiría la identificación de los lugares donde la diversidad podría reducirse como resultado de la deforestación.
- Evaluar la capacidad del modelo para predecir cambios en el uso del suelo a diferentes escalas y con el uso de diferentes conductores, tanto biofísicos como socioeconómicos.

En el área de la modelación ambiental hay cierto escepticismo a la hora de utilizar redes neuronales por considerarse un modelo de “caja negra” (Tingsanchali & Gautam, 2000). Esto se debe a que la aplicación de RNAs por lo general implica la composición de funciones no lineales que pueden ser difíciles de simplificar y reducir a términos que puedan ser comprendidos fácilmente. Lo que dificulta, por ejemplo analizar la red después de que ha sido construida y entrenada para determinar la importancia relativa de diferentes variables de entrada en la predicción de la variable dependiente (Resop, 2006). Sin embargo es importante resaltar que **las redes neuronales no deben ser consideradas una caja negra**, si bien el análisis del funcionamiento interno de la red es una tarea compleja, varias investigaciones han interpretado la contribución de las variables de entrada (Intrator & Intrator, 2001) y han proporcionado diferentes enfoques metodológicos para la interpretación de las variables (Gevrey et al. 2003) por lo tanto, más que ser un modelo predictivo podría ser usado como un modelo explicativo.

## BIBLIOGRAFÍA

1. Ali, J., Benjamin, A., Hammad, A. & Dick, O., 2005. The road to deforestation: An assessment of forest loss and its causes in Basho Valley, Northern Pakistan. *Global Environmental Change*, 15(4), pp.370 – 380.
2. Armenteras, D., Rudas, G., Rodriguez, N., Sua, S. & Romero, M., 2006. Patterns and causes of deforestation in Colombian Amazon. *Ecological Indicators*, 6, pp.353–368.
3. Bengio, Y., 1996. *Neural networks for speech and sequence recognition*, London: International Thomson Computer Press.
4. Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press.
5. Boyd, D.S., 1996. A review of “ Modelling Deforestation Processes: ” A review. Trees Series B Report nº 1. By E. F. Lambin. (Luxembourg: European Commission, 1994) [ Pp. 113“rsqb;. Free, on request. *International Journal of Remote Sensing*, 17(5), pp.1061–1062.
6. Bresee, M., Le Moine, J., Mather, S., Brososke, K., Chen, J., Crow, T. & Rademachen, J., 2004. Disturbance and landscape dynamics in the Chequamegon National Forest Wisconsin, USA, from 1972 to 2001. *Landscape Ecology*, 19(3), pp.291–309.
7. Brey, T., Jarre Teichmann & Borlich, 1996. Artificial neural network versus multiple linear regression Predicting P/B ratios from empirical data. *Marine Ecology Progress Series*, 140, pp.251–256.
8. Castejón, M., Ordieres, J.B., Vergara, E.P., Martínez-de- Pisón, F.J., Pernía, A.V. & Alba, F., 2010. AMORE: AMORE flexible neural network package. *R package version 0.2-12*. Available at: URL <http://CRAN.R-project.org/package=AMORE>.
9. Chomitz, K.M. & Gray, D.A., 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *World Bank Economic Review*, 10(3), pp.487–512.
10. Colasanti, R.L., 1991. Discussions of the possible use of neural network algorithms in ecological modelling. *Binary Computing In Microbiology*, 3(1), pp.13–15.
11. Comrie, A.C., 1997. Comparing Neural Networks and Regression Models for Ozone Forecasting. *Journal of the Air & Waste Management Association*, 47(6), pp.653–663.
12. Cropper, M., Puri, J. & Griffiths, C., 2001. Predicting the Location of Deforestation: The Role of Roads and Protected Areas in North Thailand. *Tropical Deforestation and Land Use*, 72(2), pp.172–186.

13. Esri, 2010 ArcGIS (Versión 10.1) Software de procesamiento espacial de datos. California, Estados Unidos: Environmental Systems Research Institute, Inc.
14. Etter, A., McAlpine, C., Wilson, K., Phinn, S. & Possingham., H., 2006a. Regional patterns of agricultural land use and deforestation in Colombia. *Agriculture, Ecosystems and Environment*, 114, pp.369–386.
15. Etter, A., McAlpine, C., Phinn, S., Pullar, D. & Possingham, H., 2006b. Characterizing a tropical deforestation wave: a dynamic spatial analysis of a deforestation hotspot in the Colombian Amazon. *Global Change Biology*, 12, pp.1409–1420.
16. Etter, A., McAlpine, C., Pullar D. & Possingham, H., 2006c. Modelling the conversion of Colombian lowland ecosystems since 1940: Drivers, patterns and rates. *Journal of Environmental Management*, 79, pp.74–87.
17. Ewers, R., Kliskey, A., Walker, S., Rutledge, D., Harding. & Didham, R., 2006. Past and future trajectories of forest loss in New Zealand. *Biological Conservation*, 133(3), pp.312–325.
18. Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognition Letters*. 27(8), pp.861–874.
19. Freitas, S.R., Hawbaker, T.J. & Metzger, J.P., 2010. Effects of roads, topography, and land use on forest cover dynamics in the Brazilian Atlantic Forest. *Forest Ecology and Management*, 259(3), pp.410–417.
20. Gaston, Pressey, R.L. & Margules, C.R., 2002. Persistence and vulnerability: Retaining biodiversity in the landscape and in protected areas. *Journal of Biosciences*, 27(4), pp.361–384.
21. Geist, H. & Lambin, E., 2001. *What drives tropical deforestation? A meta- analysis of proximate and underlying causes of deforestation based on subnational case study evidenc*, Belgium. CIACO Louvain-la-Neuve.
22. Gevrey, M., Dimopoulos, I. & Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Modelling the structure of aquatic communities: concepts, methods and problems.*, 160(3), pp.249–264.
23. Gobernación de Antioquia, IDEA & IGAC, 2007. *Antioquia características geográficas*, Bogotá, Colombia: Imprenta Nacional de Colombia.
24. Gómez-Sanchis, J., Martín-Guerrero, J., Soria-Olivas, E., Vila-Francés, J., Carrasco J. & del Valle Tascón, S ., 2006. Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration. *Atmospheric Environment*, 40(32), pp.6173–6180.

25. Gong, P., 1996. Integrated analysis of spatial data from multiple sources : Using evidential reasoning and artificial neural network techniques for geological mapping. *Photogrammetric Engineering & Remote Sensing*, 62(5), pp.513–523.
26. Goss, E.P. & Vozikis, G.S., 2002. Improving Health Care Organizational Management Through Neural Network Learning. *Health Care Management Science*, 5(3), pp.221–227.
27. Günther, F. & Fritsch, S., 2010. neuralnet: Training of Neural Networks. *The R Journal*, 2(1), pp.2073–4859.
28. Haibo He & Garcia, E., 2009. Learning from Imbalanced Data. *Knowledge and Data Engineering, IEEE* , 21(9), pp.1263–1284.
29. Haykin, S., 1999. *Neural Networks, A Comprehensive Foundation 2<sup>a</sup> ed.*, Hamilton, Canada: Prentice- Hall International, Inc.
30. Hecht-Nielsen, R., 1987. Kolmogorov's mapping neural network existence theorem. En Proceedings of IEEE First Annual International Conference on Neural Networks. pp. III–11.
31. Hilbert, D.W. & Ostendorf, B., 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecological Modelling*, 146(1–3), pp.311–327.
32. Hruschka, H., 1993. Determining market response functions by neural network modeling: A comparison to econometric techniques. *European Journal of Operational Research*, 66(1), pp.27–35.
33. Huston, M., 1994. *The Coexistence of Species on Changing Landscapes*, Cambridge, U.K: Cambridge University Press.
34. Intrator, O. & Intrator, N., 2001. Interpreting neural-network results: a simulation study. *Computational Statistics & Data Analysis*, 37(3), pp.373–393.
35. Jensen, J.R., Qiu, F. & Ji, M., 1999. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *International Journal of Remote Sensing*, 20(14), pp.2805–2822.
36. King, S., 2003. *Using ROC curves to compare neural networks and logistic regression for modeling individual noncatastrophic tree mortality*. General Technical Report North Central Research Station, USDA Forest Service (NC-234):3349-358.
37. Landis, R. & Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), pp.159–174.

38. Laurance, W.F., Albernaz, A.K., Schroth, G., Fearnside, P., Bergen, S., Venticing, E. & Da Costa, C., 2002. Predictors of Deforestation in the Brazilian Amazon. *Journal of Biogeography*, 29(5-6), pp.737–748.
39. Laurance, W.F., Goosem, M. & Laurance, S.G.W., 2009. Impacts of roads and linear clearings on tropical forests. *Trends in Ecology & Evolution*, 24(12), pp.659–669.
40. Lawrence, S., Giles, C. & Tsoi, A., 1996. *What size neural network gives optimal generalization? convergence properties of backpropagation*, University of Maryland College Park. Technical Report. UMIACS-TR-96-22 and CS-TR-3617
41. Lek-Ang, S., Deharveng, L. & Lek, S., 1999. Predictive models of collembolan diversity and abundance in a riparian habitat. *Ecological Modelling*, 120(2–3), pp.247–260.
42. Lek, S. & Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120(2–3), pp.65–73.
43. Li, X. & Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16(4), pp.323–343.
44. Linkie, M., Smith, R. & Leader-Williams, N., 2004. Mapping and predicting deforestation patterns in the lowlands of Sumatra. *Biodiversity & Conservation*, 13(10), pp.1809–1818.
45. Lipton, M. & Ravallion, M., 1995. Poverty and Policy. En *Handbook of Development Economics III*. New York: Elsevier: In J. Behrman and T. N. Srinivasan, eds.
46. Locatelli, B., herawati, H., Brockhaus, M., Indinoba, M. & Kanninen, M., 2008. Methods and tools for assessing the vulnerability of forests and people to climate change. CIFOR Working paper no. 43
47. López, S. & Blanco, J., 2008. Illicit Crops in Tropical America: Deforestation, Landslides, and the Terrestrial Carbon Stocks. *Ambio: A Journal of the human Environment*, 37(2), pp.141–143.
48. Mallard, F. & Francois, D., 2012. Effectiveness of the legal framework for natural areas protection relative to French road projects. *Land use policy*, 30(1), pp.582–591.
49. Manel, S., Dias, J.-M. & Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 120(2–3), pp.337–347.
50. Maier, H.R. & Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15(1), pp.101–124.

51. Mas, J.F., Puig, H., Palacio, J.L. & Sosa-López, A., 2004. Modelling deforestation using GIS and artificial neural networks. *Environmental Modelling & Software*, 19(5), pp.461 – 471.
52. McCulloch, W.S. & Pitts, W., 1943. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biology*, 5, pp.115–133.
53. Mertens, B., Kaimowitz, D., Puntode, A., Vanclay, J. & Mendez, P., 2004. Modelling Deforestation at Distinct Geographic Scales and Time Periods in Santa Cruz, Bolivia. *International Regional Science*, 27(3), pp.271–296.
54. Mertens, B. & Lambin, E., 1997. Spatial modelling of deforestation in southern Cameroon. *Applied Geography*, 17, pp.143–162.
55. METI & NASA, 2009. ASTER Global Digital Elevation Map. Obtenido de <http://asterweb.jpl.nasa.gov/gdem.asp>.
56. Meyer, W.B. & Turner, B.L., 1994. *Data on global land cover change in change in land use and land cover: a global perspective* Cambridge University Press., Cambridge England.
57. Ministerio de Ambiente, Vivienda y Desarrollo Territorial, Instituto Geográfico Agustín Codazzi (IGAC) & Conservación Internacional Colombia, 2005. *Reservas forestales protectoras nacionales, atlas básico Colombia 2ª ed.* Bogotá, Colombia: IGAC.
58. Moreira, M. & Fiesler, E., 1995. Neural Networks with Adaptive Learning Rate and Momentum Terms. IDIAP Technical Report no 95-04.
59. Müller, D. & Mburu, J., 2009. Forecasting hotspots of forest clearing in Kakamega Forest, Western Kenya. *Forest Ecology and Management*, 257, pp.968–977.
60. Nelson, G., De Pinto, A., Harris, V. & Stone, S., 2004. Land Use and Road Improvements: A Spatial Perspective. *International Regional Science Review*, 27, pp.297–325.
61. Nelson, G., Harris, V. & Stone, S., 1999. Spatial Econometric Analysis and Project Evaluation: Modeling Land Use Change in the Darién. Inter-American Development Bank.
62. Nelson, G. & Hellerstein, D., 1997. *Do roads cause deforestation? Using satellite images in econometric analysis of land use*, American Journal of Agricultural Economics, 79(1), pp.80-88.
63. Nguyen, G.H., Bouzerdoun, A. & Phung, S., 2009. Learning pattern classification tasks with imbalanced data sets. En *Pattern recognition*. Vukovar, Croatia: Intech, pp. 193–208.



64. Okwuashi, O., Isong, M., Eyo, E., Eyoh, A., Nwanekezie, O., Olayinka, D., Udoudo, D. & Ofem, B., 2012. GIS Cellular Automata Using Artificial Neural Network for Land Use Change Simulation of Lagos, Nigeria. *Journal of Geography and Geology*, 4(2), pp.94–101.
65. Orrego, S., 2009. *Economic modeling of tropical deforestation in Antioquia (Colombia), 1980-2000: an analysis at a semi-fine scale with spatially explicit data*. Ph D Dissertation. Oregon State University.
66. Overmars, K., de Koning, G. & Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164(2-3), pp.257–270.
67. Pan, Y., Roth, A., Yu, Z. & Doluschitz., 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems*, 34(5), pp.400 – 408.
68. Paruelo, J. & Tomasel, F., 1997. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecological Modelling*, 98(2–3), pp.173–186.
69. Perrone, M.P. & Cooper, L.N., 1993. *When networks disagree: ensemble methods for hybrid neural networks*, Chapman and Hall.
70. Plagianakos, V.P., Magoulas, G.D. & Vrahatis, M.N., 2001. Learning Rate Adaptation in Stochastic Gradient Descent. En N. Hadjisavvas & P. Pardalos, eds. *Advances in Convex Analysis and Global Optimization*. Nonconvex Optimization and Its Applications. Springer US, pp. 433–444. Available at: [http://dx.doi.org/10.1007/978-1-4613-0279-7\\_27](http://dx.doi.org/10.1007/978-1-4613-0279-7_27).
71. Pontius Jr, P. & Batchu, K., 2003. Using the Relative Operating Characteristic to Quantify Certainty in Prediction of Location of Land Cover Change in India. *Transactions in GIS*, (4), pp.467–484.
72. Pontius Jr., R.G., Cornell, J.D. & Hall, C.A., 2001. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Predicting Land-Use Change*, 85(1–3), pp.191–203.
73. Puyravaud, J.-P., 2003. Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, 177(1–3), pp.593–596.
74. R Development Core Team, 2013. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. URL <http://www.R-project.org/>.

75. RStudio, 2013. RStudio: Integrated development environment for R (Version 0.97.551) [Computer Software]. Boston, MA. (Available <http://www.rstudio.org/>). Accessed 2013 Jan 5.
76. Ramírez, C., Orrego, S., 2011. Modelación económica con información espacialmente explícita de la deforestación en Urabá, Colombia, 1980-2000. *Semestre Económico*, 14, n. spe29, pp. 31-51.
77. Razi, M.A. & Athappilly, K., 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29(1), pp.65–74.
78. Resop, J., 2006. *A Comparision of Artificial Neural Networks and Stadistical Regression with Biological Resources Applications*. MSc. Maryland: University of Maryland, College Park.
79. Rodríguez, N., Armenteras, D. & Alumbroeros, J.R., 2012. Land use and land cover change in the Colombian Andes: dynamics and future scenarios. *Journal of Land Use Science*, 8(2), pp.154–174.
80. Rojas, R., 1996a. Statistics and Neural Networks. En *Neural Networks*. Springer Berlin Heidelberg, pp. 227–261. Available at: [http://dx.doi.org/10.1007/978-3-642-61068-4\\_9](http://dx.doi.org/10.1007/978-3-642-61068-4_9).
81. Rojas, R., 1996b. The Backpropagation Algorithm. En *Neural Networks*. Springer Berlin Heidelberg, pp. 149–182. Available at: [http://dx.doi.org/10.1007/978-3-642-61068-4\\_7](http://dx.doi.org/10.1007/978-3-642-61068-4_7).
82. Rudel, T.K., Defries, R., Asner, G.P. & Lurance, W.F., 2009. Changing Drivers of Deforestation and New Opportunities for Conservation. *Conservation Biology*, 23(6), pp.1396–1405.
83. Rumelhart, D.E., Hinton, G. & Williams, 1986. Learning representations by back-propagating errors. *Nature*, 323, pp.533–536.
84. Sanabria, N.J., 2009. *Spatial Modelling and Prediction of Tropical Forest Conversion in the Isiboro Sécure National Park and Indigenous Territory (TIPNIS), Bolivia*. MSc. The Netherlands: International Institute For Geo-Information Science and Earth Observation.
85. Savage, A., Guillen, R., Lamilla, I. & Soto, L., 2010. Developing an effective community conservation program for cotton-top tamarins (*Saguinus oedipus*) in Colombia. *American Journal Of Primatology*, 72(5), pp.379–390.

86. Secretaría de Infraestructura Física, 2009. Antioquia Mapa vial. Republica de Colombia, Departamento de Antioquia.
87. Soares-Filho, B., Coutinho, G. & Lopez, P., 2006. Modelling conservation in the Amazon basin. *Nature*, 440, pp.520–523.
88. Thies, B., Meyer, H., Nauss, T. & Bendix, J., 2012. Projecting land-use and land-cover changes in a tropical mountain forest of Southern Ecuador. *Journal of Land Use Science*, 9(1), pp.1–33.
89. Tingsanchali, T. & Gautam, M.R., 2000. Application of tank, NAM, ARMA and neural network models to flood forecasting. *Hydrological Processes*, 14(14), pp.2473–2487.
90. Unidad Administrativa Especial del Sistema de Parques Nacionales Naturales de Colombia, 2006. Parque nacional natural y sitio de patrimonio mundial Los Katíos.
91. Veléz, N., Correa, I.C. & Ríos, G., 2010. Deforestación, ordenación forestal y campesinado. Proyecto de Ordenación Forestal Sostenible en el Bajo Cauca y Nordeste de Antioquia - Colombia. Informe técnico 75 p.
92. Warner, B. & Misra, M., 1996. Understanding Neural Networks as Statistical Tools. *The American Statistician*, 50(4), pp.284–293.
93. Wilson, D.R. & Martinez, T.R., 2003. The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16(10), pp.1429–1451.
94. Wilson, K., Newton, A., Echeverría, C., Wetson, C. & Burgman, M., 2005a. A vulnerability analysis of the temperate forests of south central Chile. *Biological Conservation*, 122(1), pp.9–21.
95. Wilson, K., Pressey, R., Newton, A., Burgman, M., Possingh, H. & Weston, C., 2005b. Measuring and Incorporating Vulnerability into Conservation Planning. *Environmental Management*, 35(5), pp.527–543.
96. Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Computers & Geosciences*, 35(6), pp.1125–1138.
97. Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, pp.159–175.

## ANEXOS

**ANEXO A.** Correlación entre las variables explicadoras de la deforestación para diferentes tamaños de muestra.

**Anexo A 1.** Matriz de correlación para un tamaño de muestra de 700.000 puntos.

	Distancia a ríos	Distancia a vías	Distancia de menor costo a las cabeceras municipales	Distancia al borde del bosque	Aptitud Agrícola	Pendiente
Distancia a ríos	1	0.3585	0.0591	0.3871	0.0314	-0.0514
Distancia a vías	0.3585	1	0.6183	0.4574	-0.0766	-0.0067
Distancia de menor costo a las cabeceras municipales	0.0591	0.6183	1	0.3556	-0.1553	0.1314
Distancia al borde del bosque	0.3871	0.4574	0.3556	1	-0.1115	-0.0010
Aptitud Agrícola	0.0314	-0.0766	-0.1553	-0.1115	1	-0.1654
Pendiente	-0.0514	-0.0067	0.1314	-0.0010	-0.1654	1

**Anexo A 2.** Matriz de correlación para un tamaño de muestra de 250.000 puntos.

	Distancia a ríos	Distancia a vías	Distancia de menor costo a las cabeceras municipales	Distancia al borde del bosque	Aptitud Agrícola	Pendiente
Distancia a ríos	1	0.3327	0.0357	0.3477	0.0752	-0.0748
Distancia a vías	0.3327	1	0.6058	0.4834	-0.1043	0.0540
Distancia de menor costo a las cabeceras municipales	0.0357	0.6058	1	0.3541	-0.2030	0.2073
Distancia al borde del bosque	0.3477	0.4834	0.3541	1	-0.1202	0.0304
Aptitud Agrícola	0.0752	-0.1043	-0.2030	-0.1202	1	-0.2189
Pendiente	-0.0748	0.0540	0.2073	0.0304	-0.2189	1

**Anexo A 3.** Matriz de correlación para un tamaño de muestra de 20.000 puntos

	Distancia a ríos	Distancia a vías	Distancia de menor costo a las cabeceras municipales	Distancia al borde del bosque	Aptitud Agrícola	Pendiente
<b>Distancia a ríos</b>	1	0,3585	0,0701	0,3933	0,0095	-0,0251
<b>Distancia a vías</b>	0,3585	1	0,621	0,4645	-0,0798	0,0106
<b>Distancia de menor costo a las cabeceras municipales</b>	0,0701	0,621	1	0,3686	-0,1612	0,1432
<b>Distancia al borde del bosque</b>	0,3933	0,4645	0,3686	1	-0,1471	0,025
<b>Aptitud Agrícola</b>	0,0095	-0,0798	-0,1612	-0,1471	1	-0,1942
<b>Pendiente</b>	-0,0251	0,0106	0,1432	0,025	-0,1942	1

**ANEXO B.** Redes entrenadas con diferente conjunto de Parámetros.

RED	Momento	Tasa de entrenamiento	Número de neuronas ocultas	Error Cuadrático medio	AUC
					Validación
1	0.5	0.01	6	0.180053	0.801365
2	0.5	0.05	6	0.182355	0.797864
3	0.9	0.1	6	0.183079	0.795651
4	0.5	0.01	12	0.177283	0.807358
5	0.5	0.05	12	0.180077	0.806704
6	0.9	0.1	12	0.179011	0.804450
7	0.5	0.01	13	0.175139	0.811266
8	0.5	0.05	13	0.176082	0.811374
9	0.9	0.1	13	0.180744	0.803198
10	0.5	0.01	15	0.175048	0.811941
11	0.5	0.05	15	0.176939	0.809138
12	0.9	0.1	15	0.180331	0.802318
13	0.5	0.01	18	0.172721	0.815379
14	0.5	0.05	18	0.175631	0.812162
15	0.9	0.1	18	0.178323	0.806855
16	0.5	0.01	20	0.173348	0.814942
17	0.5	0.05	20	0.174668	0.813932
18	0.9	0.1	20	0.176491	0.809250
19	0.5	0.01	25	0.170802	0.818888
20	0.5	0.05	25	0.172912	0.818829
21	0.9	0.1	25	0.176488	0.810141
22	0.5	0.01	30	0.170392	0.820893
23	0.5	0.05	30	0.171380	0.821269
24	0.9	0.1	30	0.174781	0.812837

## **ANEXO C .Código implementado en RStudio.**

La zona cubierta por bosque para 1980 en formato raster (pixel 30m) se convirtió a puntos, donde cada punto representa un pixel de 30m, para estos puntos se extrajo el valor de cada variable con la herramienta **Extract Multi values to points** de ArcGis 10.1, lo que permitió obtener una tabla con el valor de cada variable (columnas) para todos los puntos (filas), que luego se leyó en el software RStudio para la construcción y estimación de los modelos.

Primero se crearon varias funciones para el procesamiento de datos y la implementación del modelo de red neuronal que se explican a continuación.

### **Función para la estandarización.**

Sea **m** la matriz correspondiente a las variables explicadoras se creó la función “estándar1” para acotar los valores entre 0 y 1. Aunque este tipo de estandarización no es necesaria se realizó con el fin de mejorar las propiedades de convergencia de la red.

```
estandar1<-function (m)
{
  i<-0
  j<-0
  mincol<-0
  maxcol<-0
  for(j in 1:ncol(m))
  {
    mincol<-min([,j])
    maxcol<-max([,j])
    for(i in 1:nrow(m))
    {
      m[i,j]<-(m[i,j]-mincol)/(maxcol-mincol)
    }
  }
  m
}
```

### **Función para elegir una muestra balanceada**

Sea la matriz de entrada **m** cuya última columna corresponde a la variable deforestación (variable dependiente binaria), la función elige de forma aleatoria un subconjunto con la

cantidad especificada por el usuario de filas que contienen unos y ceros en la variable dependiente.

**nu** :numero de unos que se quieren en la matriz balanceada

**nc** :numero de ceros que se quieren en la matriz balanceada

```
balancemuestra<-function(m,nu,nc)
{
  unostempm<-m[m[,ncol(m)]==1,]
  cerostempm<-m[m[,ncol(m)]==0,]

  if(nu<=nrow(unostempm) & nc<=nrow(cerostempm))
  {unostempm2<-unostempm[sample(1:nrow(unostempm),nu,replace=F),]
  cerostempm2<-cerostempm[sample(1:nrow(cerostempm),nc,replace=F),]
  bm<-rbind(unostempm2,cerostempm2)
  bm<-bm[sample(1:nrow(bm),nrow(bm),replace=F),]
  }
  else
  {bm<-NULL
  print("error not enough ones or ceros on the input matrix to reach the wanted nu or nc
  numbers, plase reduce them or change the input matrix :)")
  }
  bm
}
```

**Función para la trasformación de valores continuos a categóricos para estimar el coeficiente de Kappa.**

**Vector:** vector columna de los valores estimados por el modelo

**Threshold:** valor del umbral

```
tobinary<-function(vector,threshold)
{tmpvector<-matrix(0,nrow(vector),1)
for(i in 1:nrow(vector))
{
  if(vector[i]>treshold)
  {tmpvector[i]<-1
  }
  else
  {tmpvector[i]<-0
  }
}
tmpvector
}
```



### **Función para estimar el coeficiente de Kappa**

Sea el vector columna simul, los valores de la deforestación estimada por la red convertidos a valores binarios para un determinado umbral, y el target el vector columna de los valores observados, se definió la siguiente función.

```
calculatekappa<-function(simul,target)
{totalobs<-nrow(simul)
  pp<-0
  nn<-0
  pn<-0
  np<-0
  for(i in 1:totalobs)
  {if(simul[i]==1)
    {if(target[i]==1)
      {pp<-pp+1
      }
    else
      {pn<-pn+1
      }
    }
  if(simul[i]==0)
  {if(target[i]==0)
    {nn<-nn+1
    }
  else
    {np<-np+1
    }
  }
  }
  tmpmatrix<-matrix(0,2,2)
  tmpmatrix[1,1]<-pp
  tmpmatrix[1,2]<-pn
  tmpmatrix[2,1]<-np
  tmpmatrix[2,2]<-nn
  tmpmatrix
  pcero<-(pp+nn)/totalobs
  pe1<-(pp+pn)*(pp+np)
  pe2<-(nn+np)*(nn+pn)
  pe<-(pe1+pe2)/(totalobs)^2
  kappa<-(pcero-pe)/(1-pe)
  kappa
}
```

Calculo de la matriz de confusión

Matriz de confusión

Coeficiente de Kappa

## Entrenamiento de la red neuronal

Se utilizo el paquete AMORE el cual hace uso de la función newff la cual crea un perceptron multicapa unidireccional con el método de gradiente decreciente con momento (ADAPTgdwm). Este método hace uso del algoritmo de retropropación en modo on line. Para la capa oculta y la capa de salida se utilizo la función **sigmoid** la cual hace referencia a la función de activación logística.

**P:** matriz de entrenamiento (250.000 datos)

**Target:** vector columna de la matriz de entrenamiento correspondiente a la variable dependiente.

**Base estándar:** matriz de datos total (7.758.886 datos)

require(AMORE)

```
net.start <- newff(n.neurons=c(6,12,1),
  learning.rate.global=0.05,
  momentum.global=0.5,
  error.criterium="LMS",
  hidden.layer="sigmoid",
  output.layer="sigmoid",
  method="ADAPTgdwm")
result <- train(net.start, P,target, error.criterium="LMS", report=TRUE,
show.step=100,n.shows=50)
```

Luego de entrenada la red se realiza la simulación con el conjunto de datos total

```
simul<-sim.MLPnet(result$net,base_estandar)
```

## Validación: curva ROC/AUC

Luego de simular con la redes entrenadas se procedió a evaluar el área bajo la curva ROC para cada una de ellas, para lo cual se utilizo el paquete ROCR.

**Simul:** vector columna de los valores generados por la simulación

**Target :** vector columna de la variable dependiente.

```
require (ROCR)
pred<-prediction(simul,target)
perf<-performance(pred,"tpr", "fpr")
plot(perf)
auc <- performance(pred,"auc")
auc <- unlist(slot(auc, "y.values"))
```

## Ensamble de Redes Neuronales

Luego de evaluar el área bajo la curva ROC (AUC) de las redes entrenadas, se realizó el ensamble con las redes que obtuvieron mayores valores de AUC, para lo cual se utilizó el siguiente código.

```
i<-0
comite_simul<-matrix(0,nrow(red_16),1)
for (i in 1:nrow(red_16))
{comite_simul[i]<-(red_16[i]+red_19[i]+red_20[i]+red_22[i]+red_23[i])/5}
}
```

Los valores obtenidos del comité se transformaron a valores categóricos para diferentes umbrales con la función “**tobinary**” y luego se estimó el coeficiente de Kappa para cada umbral con la función “**calculatekappa**”.

## Generación de Mapas

Se utilizó el paquete **foreign** para leer el **.dbf** correspondiente al shape de puntos (base) de las zonas en bosque para 1980. Cada punto representa un pixel de 30m y contiene el valor de las variables explicadoras y de la variable dependiente (defor) en la última columna, los valores de ésta columna se remplazaron por los valores estimados del ensamble de redes utilizando el siguiente código.

```
require(foreign)
t<-read.dbf("base.dbf")
v<-t$defor<-c(comite_simul)
write.dbf(v,"base.dbf")
```

Una vez se remplazaron los valores y el **.dbf** asociado al shape de puntos, en ArGIS 10.1 los puntos se convirtieron a raster utilizando la herramienta **Conversión Tools>To raster>Point to raster** y especificando la columna que contiene el valor estimado por el ensamble de redes. Este mismo procedimiento se realizó para generar el mapa del escenario simulado.

### **Modelo de Regresión logística**

Para seleccionar la muestra de forma aleatoria (700.000) se creó la siguiente función

```
randomsample<-function(m,rows)
{
  tmpvector<-sample(1:nrow(m),rows,replace=F)
  muestra<-matrix(0,rows,ncol(m))
  muestra<-matrix(as.matrix(m[tmpvector,1:ncol(m)]),rows,ncol(m))
  muestra
}
```

### **Estimación del modelo de regresión logística.**

Se utilizó la función glm para ajustar el modelo a los datos y la función de vinculo logit.

```
modelo_1<- glm(formula=defor ~
dis_rios+dis_vias+dis_bor+apa+pend+dis_cos,family=binomial(link="logit"),
data=muestra_700mil)
summary(modelo_1)
simul <- predict(modelo_1, newdata =datos, type = "response")
```

Al igual que para el modelo de red neuronal se estimó el área bajo la curva ROC como se indico anteriormente.

### **Comparación de modelos: Grafica curvas ROC.**

```
require (ROCR)
pred<-prediction(comite_1,target)
perf<-performance(pred,"tpr","fpr")
plot(perf,col="blue")
par(new=TRUE)
pred<-prediction(logi,target)
perf<-performance(pred,"tpr","fpr")
plot(perf,col="MediumPurple")
par(new=TRUE)
abline(a=0,b=1)
```